NAT'L INST. OF STAND & TECH R.I.C.
A11103 777737

U.S. Department of Commerce
Technology Administration
National Institute of Standards and Technology

NISTIR 4843

REFERENCE

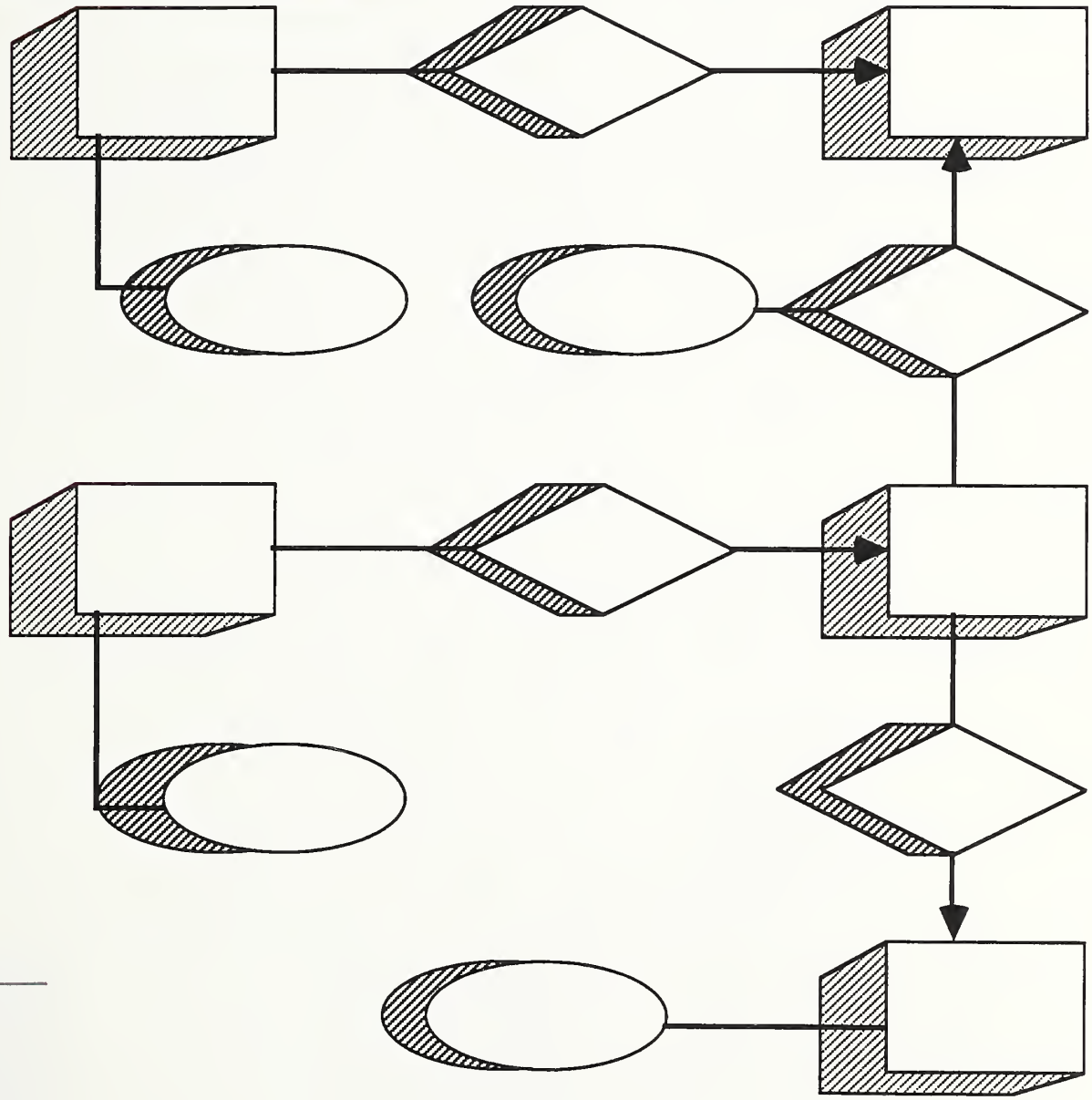
NIST
PUBLICATIONS

Managing Data: From Vision to Reality

Proceedings of the Fourth Annual DAMA Symposium

May 14-15, 1991

Judith J. Newton,
Mary Lou Melley,
Holly Higgins, Editors



QC
100
.U56
4843
1992

NISTIR
QC 100
.U56
4843
1992

NISTIR 4843

Managing Data: From Vision to Reality

Proceedings of the Fourth Annual DAMA Symposium

Judith J. Newton
Mary Lou Melley
Holly Higgins
Editors

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Gaithersburg, MD 20899

Sponsored by
Data Administration Management Association
National Capital Region

May 14 - 15, 1991

Issued May 1992

U.S. DEPARTMENT OF COMMERCE
Barbara Hackman Franklin, Secretary

TECHNOLOGY ADMINISTRATION
Robert M. White, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
John W. Lyons, Director

Table of Contents

Foreword iii

Acknowledgements v

Achieving the Data Management Reality 1
Arnold Barnett

The Road Map for CASE Implementation 7
Ian Palmer

The Impact of Implementing a New Information Architecture . . 21
Ron Shelby

Data Planning - What Good is it Any How? 45
Fred L. Forman

Data Modeling: The Data Administrator Perspective vs. the Traditional Application Development Point of View 61
Joseph Oates

Data Management and Information Proficiency: A Vision for the Future 85
Thomas Buckholtz

Information Reengineering 91
William Durell

Managing Information Across Multiple CASE Tools 101
Graham W. Thompson

The Atomic Database: Building the Perfect Beast 103
Bill Inmon

Data Management and Customer Satisfaction 109
Karen Lindsay

Data Administration, the IBM Repository, and CASE Technology at Depository Trust Company 119
Emmanuel Ackerman

Status and Application of Standards for Data Administration 127
Bruce K. Rosen

Future Vision: An Essential Prerequisite for a Data Management Program 143
Reed Phillips

Foreword

The National Capital Region chapter of the Data Administration Management Association (DAMA) sponsored its fourth annual symposium on May 14-15, 1991. The setting was the National Institute of Standards and Technology in Gaithersburg, Maryland. Several hundred attendees from the public and private sectors were treated to the wisdom of data administration practitioners and preachers at the leading edge. The topic of the 1991 symposium, "Managing Data: From Vision to Reality," set the framework for an enriching two days.

Informative presentations ranged from the practical hands-on advice to plan and implement a strong data management program offered by the first speaker, Arnold Barnett of Barnett Data Systems, to the concluding talk of Reed Phillips, director for Information Resources Management at the U.S. Department of Commerce, which provided a prescription for the vision needed to build such a program. Speakers from a variety of disciplines shared their experiences and offered advice on a wide range of data management topics: implementation of Computer-Aided Software Engineering (CASE), value-added data modeling, the global data administrator perspective, information proficiency through data management, information reengineering, sharing information via the data repository, implementation of an architected approach to data, customer satisfaction as the thrust of data management, use of the IBM repository, and an overview of data administration standards.

In addition to two days of lectures, Ronald V. Shelby, President of DAMA International, presented the 1991 DAMA-National Capital Region Achievement Award to Rebecca Wade of the Department of the Navy. Ms. Wade established data administration as a standard operating procedure for Automated Data Processing at Navy, evaluated many repository products and implemented a Department of Navy-wide data architecture. Also recognized was Kathy Hirsh of the American Management Systems for her unstinting support of this DAMA chapter since 1986.

The papers in this proceedings highlight the actual presentations. This publication is meant to be a summary record and reference for data management practitioners. Any omission does not imply criticism by DAMA or NIST. The mention of a particular company or product does not imply either endorsement or criticism by DAMA or NIST.

Acknowledgements

The 1991 DAMA Symposium was made possible through the efforts of many. They are acknowledged below. DAMA-NCR extends special thanks to the speakers who devoted their time and shared their expertise with others in the field of data administration.

Board of Directors, DAMA-NCR:

Andrea Tyndall Norris
President

Drew Falk
Vice President, Membership

Mary Lou Melley
Vice President, Administration

George Burke
Treasurer

Brad Shelton
Program Director

Mary Jane Frances Smith
Publicity Director

Daniel Prengaman
Facilities Liaison

Judith Newton
Special Advisor

Carla von Bernewitz
Newsletter Chair

DAMA Proceedings Committee:

Mary Lou Melley
Environmental Protection Agency

Judith Newton, National Institute of Standards and Technology

Brad Holoman
Vector Research Inc.

Holly Higgins, U.S. Patent and Trademark Office

Ruta Godwin
Office of Management and Budget

Marie Korn
UNISYS

Arnold Barnett

Topic: Achieving the Data Management Reality

Description: Too often data management visions never become reality. This is because of over-ambitious visions relative to resources and time and/or lack of management commitment, tools, methodology or trained personnel. Mr. Barnett will address the proper planning effort that needs to be done to determine realistic visions and to achieve the desired reality.

Biography: Mr. Barnett is the President of Barnett Data Systems of Rockville, Maryland, a 23-year old firm specializing in data administration related seminars, conferences, and consulting. He taught accounting at Penn State; worked for IBM as an accountant, systems engineer, and marketing representative; marketing representative for Control Data Corporation; and partner in Kornbluh-Barnett Associates. Mr. Barnett has an MS degree in accounting from Penn State and is a CPA, CDP, and a Certified Data Educator.

What we are trying to achieve via Data Management:

Standard Data Elements, e.g., all dates are "YYYYMMDD."

Standard Data Coding, e.g., "TX" is the standard two-character code for the state of Texas.

Standard Identifiers for Entity Items, e.g., the standard identifier for the customer XYZ Company is "423078."

Standard Physical Data names, e.g., the standard DB2 name for Social Security number of Employee is "EMPSSNIDN."

Better Database Design, i.e., stability and flexibility.

Determine all of the organization's business activities, e.g., a functional decomposition.

Determine which things the organization collects or should collect data about (entities), the relationships between the entities and their associated business rules, e.g., entity-relationship models with appropriate annotation.

Data Sharing, e.g., A subject database supports three applications from the same tables.

Data Accessibility, e.g., Data easily available to an end user on his/her micro/terminal from the local- or wide-area networks and the mainframe. Also, it implies user friendly query facilities.

Knowledge of What Data is Available, its characteristics and how it may be accessed, e.g., a data directory or locator system.

Planned Data Redundancy, e.g., minimum redundancy necessary to maintain performance to physically locate the data to enhance user access.

Data Security, e.g., protecting data against intentional and unintentional modification, deletion or theft.

A Data Environment that Facilitates Systems Development, e.g., Shared or subject databases that contain all or most of the data needed by new systems.

Factors that limit our achievements:

Lack of Proper Hardware, e.g., slow micro platforms for CASE software.

Lack of Proper Software, e.g., no data dictionary.

Lack of Knowledgeable Personnel - no experience, no training, etc.

Lack of Funds, e.g., no or limited training budget.
Lack of Time, i.e., too much to do with too few people within a given time frame.

Lack of executive Management Commitment, e.g., refusal to enforce data coding standards.

Lack of Peer Support, e.g., Systems Development management opposing data driven design.

Non-Data Driven Systems Development Methodologies or Lack of any Methodology - "business as usual".

Poor Utilization of CASE Tools, e.g., no accompanying methodology.

Unrealistic Visions of What Can be Accomplished - the vision much greater than what can be achieved with the committed resources, time, etc.

What planning activities are needed to achieve cost/beneficial, data management achievements:

Obtain commitment from a sponsor to determine the state of data management practices and recommend a plan to enhance data management in the future. Figure on two persons spending approximately thirteen weeks to accomplish the plan.

1. Conduct a Problem Definition

   Determine the current and future (business) operational problems caused by the current or planned, if any, inadequate or lacking data management practices.

   Determine the current and future data problems caused by the current or planned, if any, inadequate or lacking data management practices.

   Determine the operational and systems opportunities afforded by proper data management practices.

   Most of the problems and opportunities elicited are already known to the Data Administrator; however, it is very beneficial to elicit them from the users, executives, systems developers, data base administrators, EDP auditors, quality assurance personnel, etc.

   This "formal" statement of the problem will help get the attention of executive and peer management and make them more receptive to the "solutions" that proper data management practices will provide.

Review the results of the Problem Definition with the sponsor and determine which data management practices to establish/improve.

2. Design the data management activities that cost/beneficially address one or more items on the Problem Definition.

   Attempt to sell executive and peer management on the benefits of the plan to introduce, redirect or enhance the organization's data management activities. Any portion of the plan not sold should be removed from the design.

   This design should be reviewed on an iterative basis by the sponsor and those who will need to support and/or be beneficiaries of the data management activity when it is operational.

   This design should be cognizant of any constraints, e.g., time, personnel, money, commitment, etc. This will keep the plan realistic.

3. The design must contain:

   What data management activities will be initially performed and their scope, i.e., standard definitions will be created for all data elements used in new financial systems development efforts.

   Follow-on activities, if any, i.e., standard definitions will be created for all data elements used in any new systems development efforts. Note that the creation of standard definitions for data elements not used in new applications would be an additional follow-on effort.

   Any new hardware or software, i.e., two fast micro computers with OS/2 and a CASE tool.

   The metadata to be stored and utilized, e.g., the meta entities, the meta attributes and their relationships. In effect, the structures of the metadata base(s) to be used to effect the data management objective.

   Retrofitting plans, if any, e.g., collecting and storing metadata relative to the current data and/or process environments.

   Major policies and procedures, i.e., which policies and procedures are essential to achieving success. For example, to effect data standardization across organizational lines there must be a Corporate data standardization policy with Corporate enforcement procedures.

Staffing requirements, e.g., two additional data analysts, each with a certain level of knowledge and experience.

Timetable, i.e., a Gantt chart showing development activities and milestones and the data management operational milestones.

Costs, e.g., new hardware, software, personnel, travel and training, etc.

Benefits. These should be stated in relation to specific items in the Problem Definition. For example, the use of data modeling envisioned in the plan is in response to the current problem of unstable and inflexible databases.

4. The Problem Definition and the above planning should take approximately three months. Once approved, there will be a period of time necessary to detail the design, hire personnel, perform training, software installation, etc.

5. When everything is in place, begin the data management activity.

In summary:

First, to achieve a (cost/beneficial) data management reality, one must first ascertain the problems to be solved and the opportunities afforded by data management techniques.

Second, develop a realistic plan to introduce, redirect or enhance data management and sell it to all cognizant personnel.

Third, obtain approval to proceed.

Fourth, execute the plan by building the necessary infrastructure to successfully realize the data management objectives.

Fifth, begin operations.

Ian Palmer

Topic: The Road Map for CASE Implementation

Description: Successful implementation of Computer-Aided Software Engineering (CASE) supported by methodology throughout an enterprise is a major project typically lasting several years. Based on its practical experience over the past 12 years, James Martin Associates (JMA) has developed a six-phased approach with five task streams, to provide a sound structure for realistic management of this complex project.

The road map provides a variety of pathways, but generally begins with the careful selection of a path-finder project and leads through the planned fan-out of CASE workstations backed by the necessary infrastructure development and organizational adjustments. It explicitly recognizes the cultural change implications that determine how quickly and smoothly an information system organization can begin to achieve the benefits of CASE and methodology.

Biography: Ian Palmer is Chairman of James Martin Associates in North America and CEO of the Products division worldwide. He has been responsible for spreading the successful use of Information Engineering (IE) throughout American corporations and government agencies. His current emphasis is on development JMA's IE and AD/Cycle products including metrification, hypermedia, computer-based training, and rule-based systems.

Mr. Palmer is internationally recognized as a leading authority on IE, data base technology and systems development methodologies. He was one of the pioneers in data base, and produced the first text book on the subject, "Data Base Systems: A Practical Reference," in 1973. With British Aerospace, he led what is believed to be the first commercial use of entity relationship modeling, also in 1973. For many years, he served on the Codasyl Committee, one of the international focal points for the development of the data sharing concept.

Mr. Palmer is a Master's graduate from the University of New Zealand.

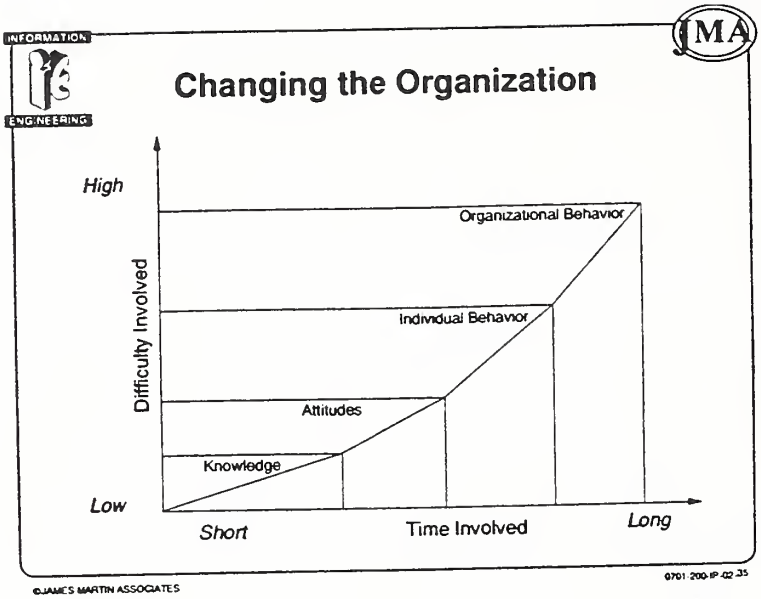
JMA

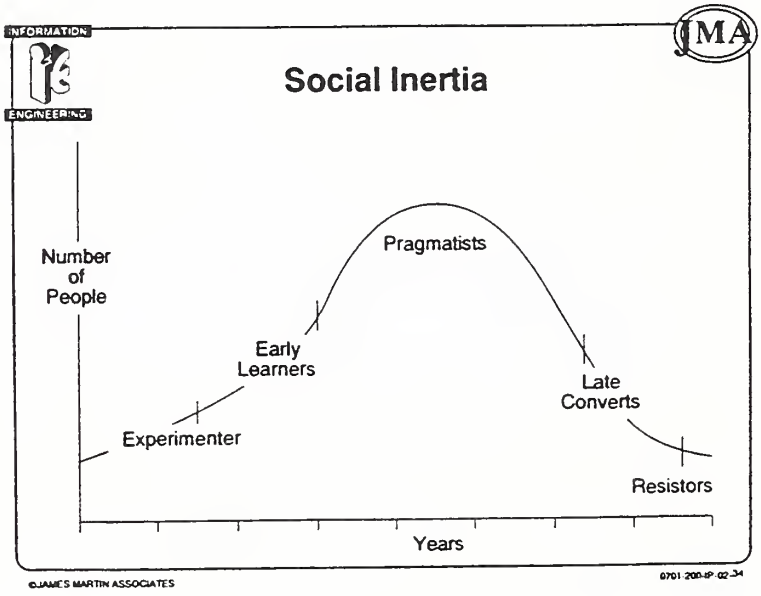
The Road Map for CASE Implementation

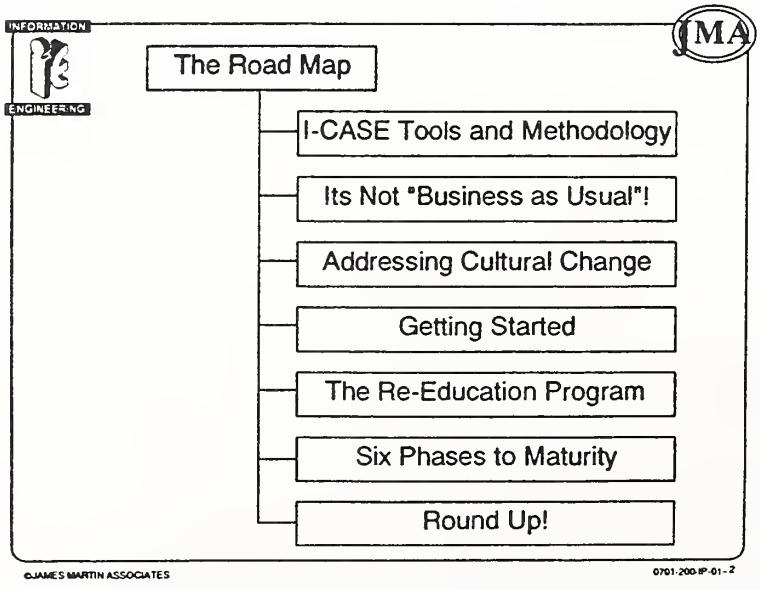
Ian Palmer
Chairman
James Martin Associates, Inc.

DAMA
Gaithersburg
May 1991

©JAMES MARTIN ASSOCIATES


INFORMATION ENGINEERING
JMA
Changing the Organization
High
Low
Difficulty Involved
Organizational Behavior
Individual Behavior
Attitudes
Knowledge
Short
Time Involved
Long
©JAMES MARTIN ASSOCIATES
0701-200-IP-02.35


INFORMATION ENGINEERING
JMA
Social Inertia
Number of People
Experimenter
Early Learners
Pragmatists
Late Converts
Resistors
Years
©JAMES MARTIN ASSOCIATES
0701-200-IP-02.34

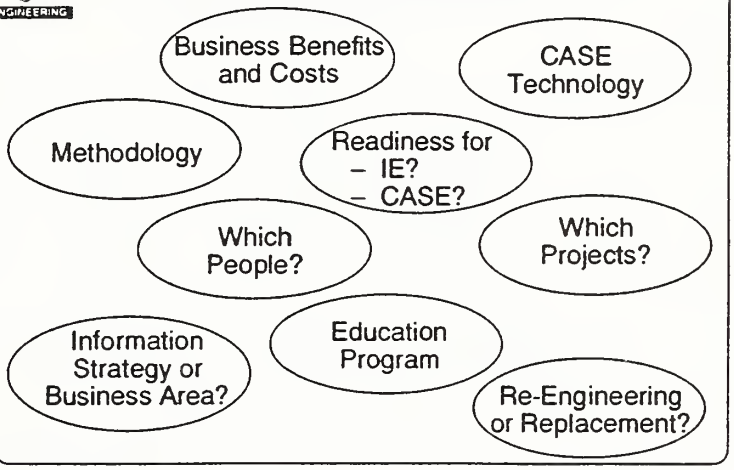

INFORMATION ENGINEERING
JMA
The Road Map
I-CASE Tools and Methodology
Its Not "Business as Usual"!
Addressing Cultural Change
Getting Started
The Re-Education Program
Six Phases to Maturity
Round Up!
©JAMES MARTIN ASSOCIATES
0701-200-IP-01-2

Some Approaches to CASE and Methodology

1. The ostrich strategy
2. The agnostic strategy
3. The piecemeal strategy
4. The pilot strategy
5. The committed management strategy
6. The big bang strategy

©JAMES MARTIN ASSOCIATES

Elements of the Strategy



©JAMES MARTIN ASSOCIATES

Pleas from the Systems Development Manager

"I need to be able to show my boss the benefits"

"We have no time for analysis and design"

"The users are not willing to participate"

"My people don't want to change"

"How do I separate CASE hype from reality?"

©JAMES MARTIN ASSOCIATES

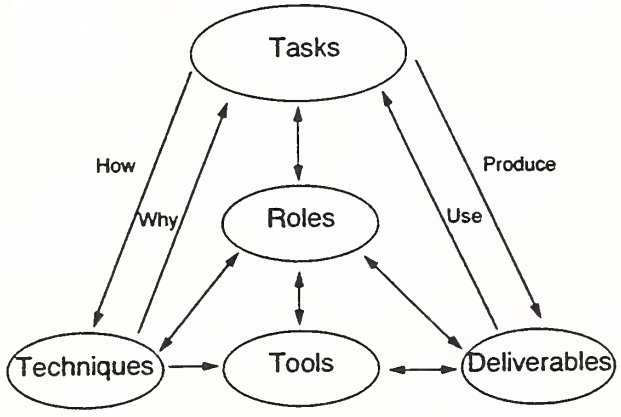
Key Features of an Integrated CASE Product

I-CASE

- Sharing an encyclopedia containing all system development objects associations
- Supporting all essential data oriented and process oriented techniques
- With a unified user interface
- Designed to enable code generation
- Used routinely by many analysts and designers

©JAMES MARTIN ASSOCIATES

Components of a Methodology



©JAMES MARTIN ASSOCIATES

Methodology is Not Merely–

- Tool Documentation
- Diagramming Standards
- Adhoc Techniques
- An Optional Extra

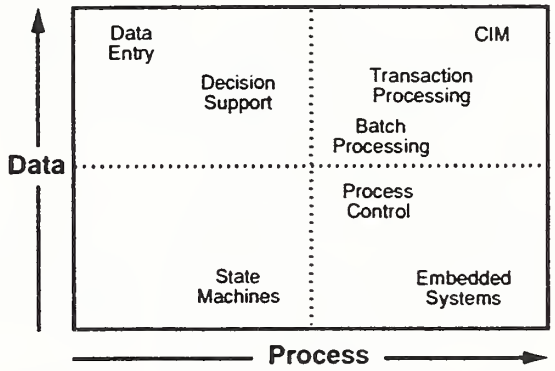
©JAMES MARTIN ASSOCIATES

Decreasing Stability

- Strategic objectives
- Business functions
- Data structures
- Processes
- Enterprise structure
- User Procedures
- Technology
- Information needs

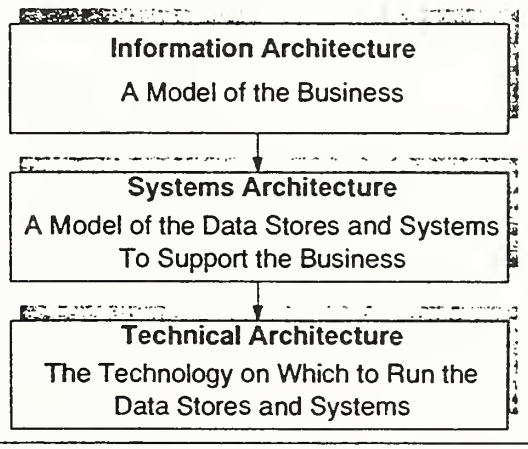
©JAMES MARTIN ASSOCIATES

Emphasis on Data and Process

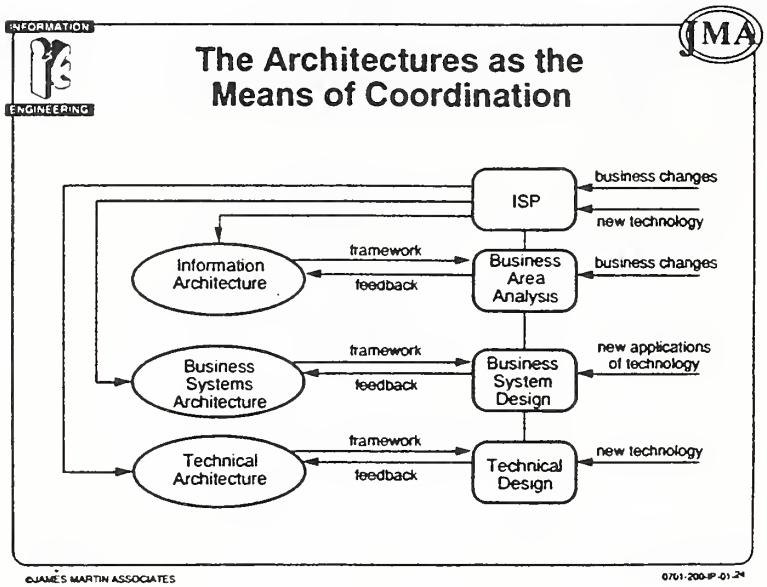


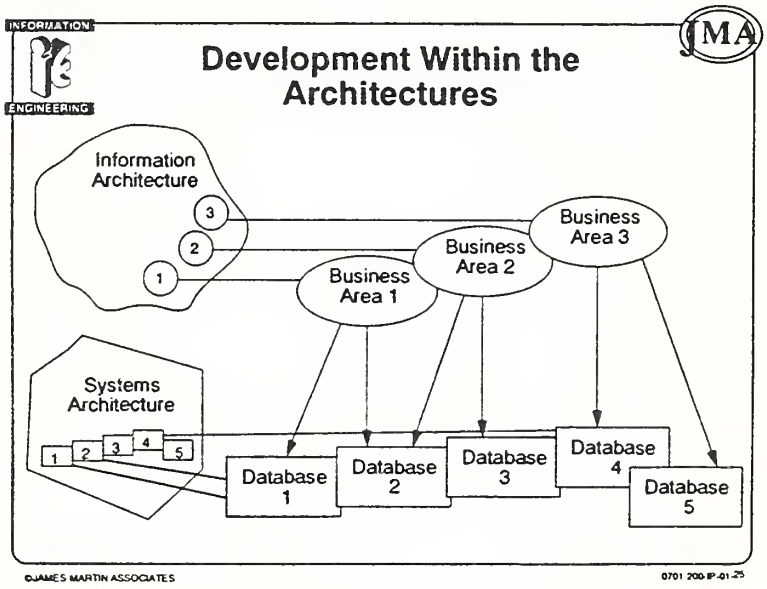
©JAMES MARTIN ASSOCIATES

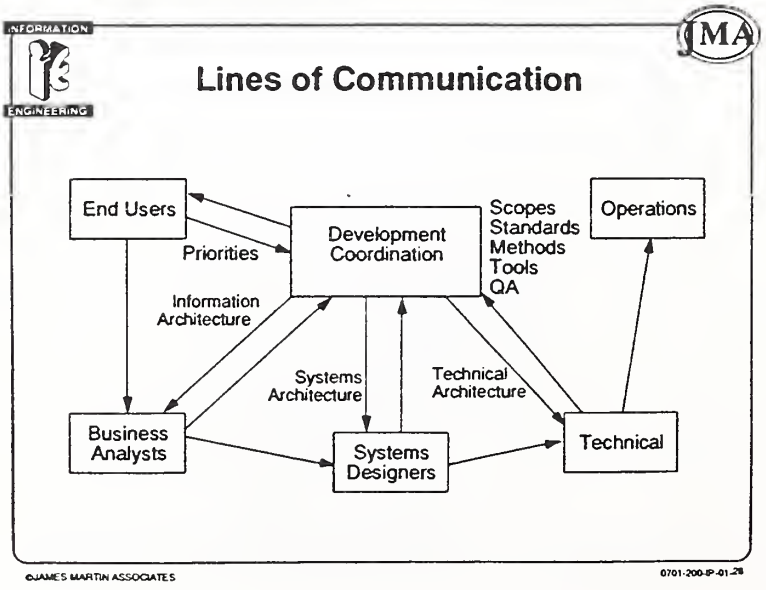
Information Engineering Architectures


Information Architecture
A Model of the Business
Systems Architecture
A Model of the Data Stores and Systems To Support the Business
Technical Architecture
The Technology on Which to Run the Data Stores and Systems

©JAMES MARTIN ASSOCIATES


INFORMATION ENGINEERING
JMA
The Architectures as the Means of Coordination
ISP
business changes
new technology
Information Architecture
framework
feedback
Business Area Analysis
business changes
Business Systems Architecture
framework
feedback
Business System Design
new applications of technology
Technical Architecture
framework
feedback
Technical Design
new technology
©JAMES MARTIN ASSOCIATES
0701-200-IP-01-24


INFORMATION ENGINEERING
JMA
Development Within the Architectures
Information Architecture
1
2
3
Business Area 1
Business Area 2
Business Area 3
Systems Architecture
1
2
3
4
5
Database 1
Database 2
Database 3
Database 4
Database 5
©JAMES MARTIN ASSOCIATES
0701-200-IP-01-25


INFORMATION ENGINEERING
JMA
Lines of Communication
End Users
Priorities
Development Coordination
Scopes
Standards
Methods
Tools
QA
Operations
Information Architecture
Systems Architecture
Technical Architecture
Business Analysts
Systems Designers
Technical
©JAMES MARTIN ASSOCIATES
0701-200-IP-01-28

The Marriage of Cultures

Information	Engineering
• Flexibility	• Rigor
• High pressure	• Metrification
• Adhoc	• Repeatability
• User oriented	• Auditabiltiy
	• Objectivity

©JAMES MARTIN ASSOCIATES

Cultural Clashes With IE

"Get it right first time."
- Iteration

"Bias for action."
- Architectures

"Perpetuate the status quo."
- Search for opportunities

"Information is power."
- Data sharing

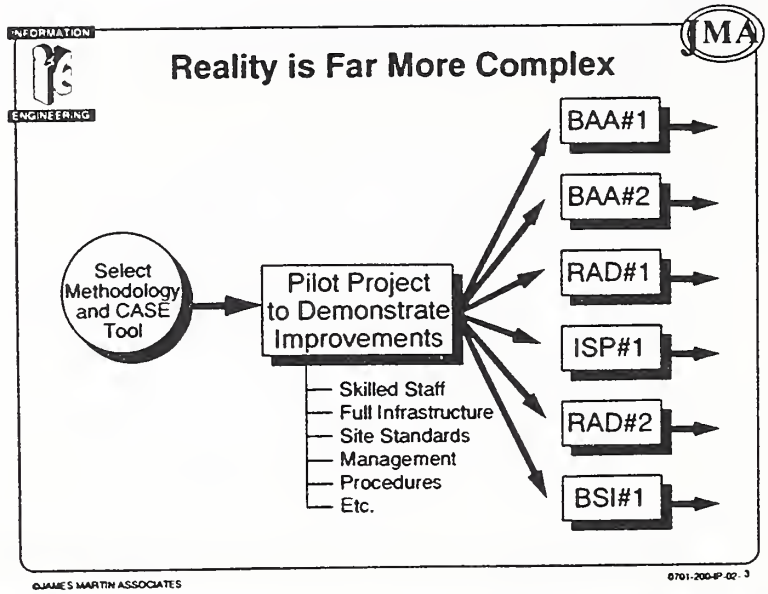
"Data versus process."
- Interaction analysis

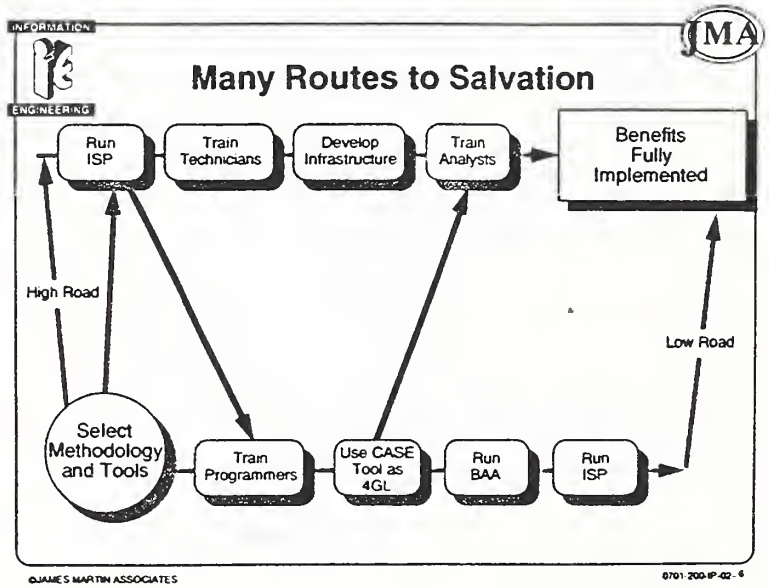
©JAMES MARTIN ASSOCIATES

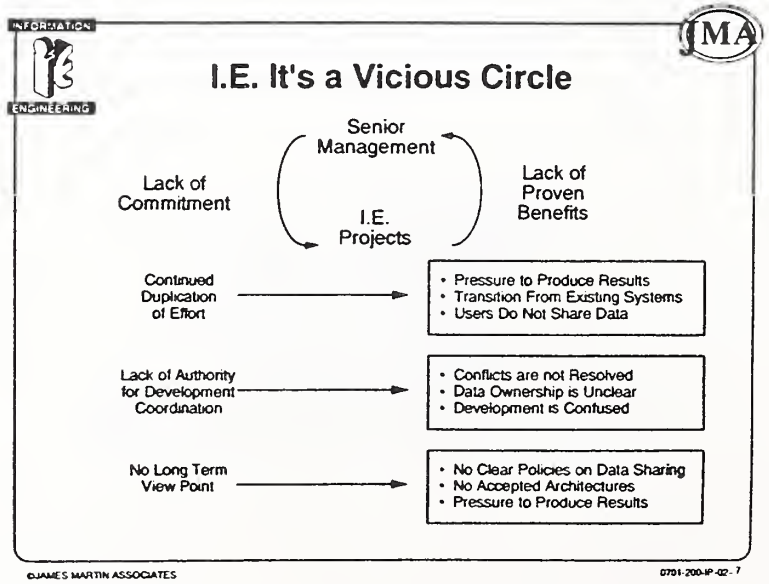
To Manage the Cultural Issues

- Recognition
- Organizational evolution
- Individual career planning
- Education and training
- Objectives and performance measures
- Pilot projects to quantify the benefits

©JAMES MARTIN ASSOCIATES


INFORMATION ENGINEERING
JMA
Reality is Far More Complex
Select Methodology and CASE Tool
Pilot Project to Demonstrate Improvements
Skilled Staff
Full Infrastructure
Site Standards
Management
Procedures
Etc.
BAA#1
BAA#2
RAD#1
ISP#1
RAD#2
BSI#1
©JAMES MARTIN ASSOCIATES
0701-200-IP-02 - 3


INFORMATION ENGINEERING
JMA
Many Routes to Salvation
Run ISP
Train Technicians
Develop Infrastructure
Train Analysts
Benefits Fully Implemented
High Road
Low Road
Select Methodology and Tools
Train Programmers
Use CASE Tool as 4GL
Run BAA
Run ISP
©JAMES MARTIN ASSOCIATES
0701-200-IP-02 - 6


INFORMATION ENGINEERING
JMA
I.E. It's a Vicious Circle
Senior Management
Lack of Commitment
Lack of Proven Benefits
I.E. Projects
Continued Duplication of Effort
• Pressure to Produce Results
• Transition From Existing Systems
• Users Do Not Share Data
Lack of Authority for Development Coordination
• Conflicts are not Resolved
• Data Ownership is Unclear
• Development is Confused
No Long Term View Point
• No Clear Policies on Data Sharing
• No Accepted Architectures
• Pressure to Produce Results
©JAMES MARTIN ASSOCIATES
0701-200-IP-02 - 7

Key Re-Education Topics

- Methodology
- Development coordination
- Quality assurance
- Project management
- RAD techniques
- Business area analysis
- Information strategy planning
- Re-engineering
- Effective use of the CASE tools
- Encyclopedia administration

©JAMES MARTIN ASSOCIATES

Participation in an Information Engineering Education Program

	Management	Users	Analysis	Designers	Programmers	Development Coordination
Understanding the Cultural Change	O	O				
Information Engineering	O	O	I	I	I	I
Information Strategy Planning	O	O				W
Business Area Analysis		W	W	I		W
Advanced Business Area Analysis			W			W
User System Design		I	W	W		W
Prototyping		W		W		I
Data Design for IMS				W	I	W
Data Design for DB2				W	I	W
Software Design and Generation				W	W	
Development Coordination	O	O	I	I	I	W
Project Management in IE			I	I	I	W
Quality assurance						W
Using the Information Center		W	W			
Using the Encyclopedia		W	W	W	W	W

O = Overview
I = Introduction
W = Workshop

©JAMES MARTIN ASSOCIATES

The IE Education Program Critical Success Factors

- Methodology before tools
- A little at a time
- Not limited to classroom training
- Coordination for immediate application
- Integrated with career development

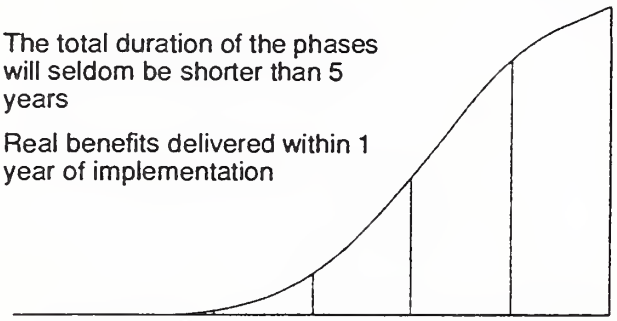
©JAMES MARTIN ASSOCIATES

INFORMATION ENGINEERING

JMA

Timescale of Phases

- The total duration of the phases will seldom be shorter than 5 years
- Real benefits delivered within 1 year of implementation



©JAMES MARTIN ASSOCIATES

INFORMATION ENGINEERING

JMA

Exploration Phase

- The internal sale
- Must be managed, but time period is imprecise
- Spread awareness over several business units
- Demonstrate the potential and implication by a small training project
- Meet successful enterprises
- Plan for cultural change
- Manage expectations

©JAMES MARTIN ASSOCIATES

INFORMATION ENGINEERING

JMA

Preparation Phase

- Customize the methodology
- Appoint program manager and stream managers
- Set up IE program infrastructure
- Top management commitment
- Promote the IE program
- Use a time box approach

©JAMES MARTIN ASSOCIATES

Initiation Phase

- Stand alone projects using IE Pathfinder techniques
- Interference from sceptics on the usage of IE
- Experimentation
- Methodology starter standards
- No immediate impact on organization
- Insignificant impact on resources or management
- But essential to get right

©JAMES MARTIN ASSOCIATES

Extension Phase

- Several development projects using IE
- Rush of the innovators to join IE projects
- IE is spreading across the enterprise
- Rapidly evolving standards
- Development coordination is established
- Has some impact on resources and skills
- Traditionalists may still kill it

©JAMES MARTIN ASSOCIATES

Integration Phase

- Not using IE in development is an exception
- Enterprise is not resisting IE
- Development coordination is formal requirement
- Organizational structure has changed as a result of IE
- IE is customized to the enterprise
- There is no way back to the old methods

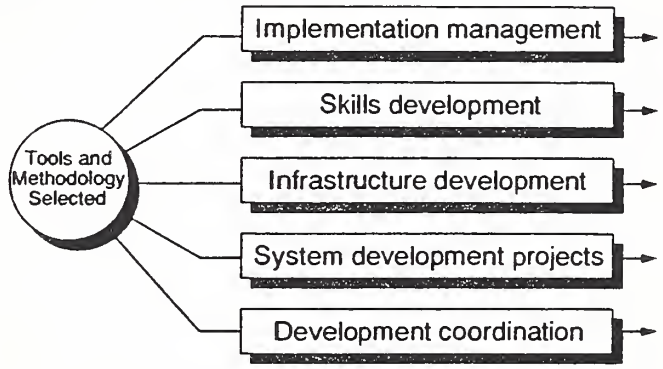
©JAMES MARTIN ASSOCIATES

Maturity Phase

- Almost all systems and databases implements with IE
- Few remaining non-IE systems controlled with IE business models
- Widespread use of information center
- Standards are embedded in IE
- Significant drop in maintenance cost
- Enterprise uses for business opportunities
- Enterprise believes in IE

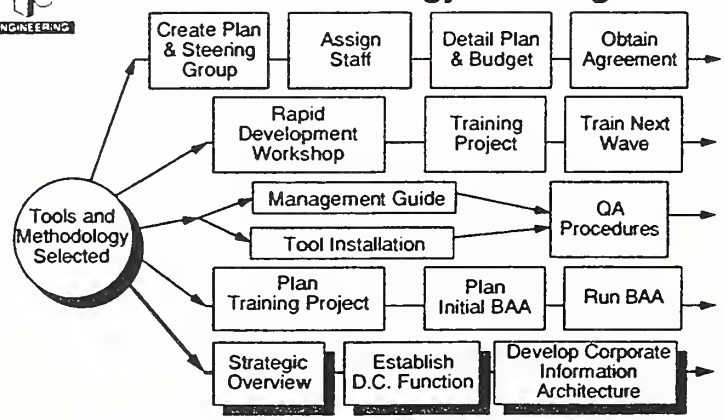
©JAMES MARTIN ASSOCIATES

Detailed I.E. Implementation Plan Before Initiation Phase



©JAMES MARTIN ASSOCIATES

Information Strategy Planning



©JAMES MARTIN ASSOCIATES

Ron Shelby

Topic: The Impact of Implementing a new Information Architecture

Description: This presentation, based upon experience with large organizations, will highlight the role of information-driven planning, analysis, and system development techniques in organizations which use information technology to cut costs, improve service, and react to changes in their environment. Ron will portray the impact of implementing an information architecture which shares data upon an organization's work process and structure. He will discuss the dimensions of these impacts, and what information resource management professionals can do to help an organization deal positively with the changes which are likely to occur.

Biography: Ron Shelby is head of Connecticut Mutual's Information and Technology Services Group, which provides data administration, information engineering, data base management, and project management services to the companies of the Connecticut Mutual Alliance. Prior to joining Connecticut Mutual he was data administrator at Travelers Canada and the US Department of Interior before becoming a consultant specializing in information resource management. Ron is Past-President of DAMA International.

What I am about to outline for you is a personal perspective on information architecture. I will discuss why it is important for an enterprise in the information age to implement a new information architecture, Connecticut Mutual's approach to this very difficult exercise and the impact it is having.

Connecticut Mutual provides life insurance and pension management services to clients throughout the United States. Our life insurance business is the seventeenth largest in the United States. Total corporate revenues exceed $3 billion and our assets exceed $11 billion. Currently, we have between 4,000-5,000 employees, and somewhere between 400-500 information systems professionals. Two years ago we decentralized the IS application development function and many individuals who perform IS work full time are no longer categorized in a way that is easy to discern that they are now full time IS professionals. Our central information systems shop, in which I am a manager, is approximately 190 individuals. Our name is Architecture, Planning and Operations division, and we provide the services described in our title, including information architecture planning support, project management and information engineering support, and data base services. The division also supports the network and operations functions for the company and our offices. Connecticut Mutual has operated a wide area network to our field offices for the past 16 years and we have more than 2 dozen LANs in service at the home and field offices.

Let me turn now to the issue of, "What is information architecture, anyway?". My own definition of information architecture in that it is a view of the information related components of an enterprise which communicates the role of each component in operating the enterprise, just as an architect's drawings and plans specify the place of each building and implicit in such a drawing is the function or role of each of these buildings or components of a building. An information architecture must convey the meaning and role of each of the subcomponents within the architecture. If you were to categorize and describe completely the information architecture of your organization today as it currently exists, you would probably be astounded at the lack of coherency.

The reason that current information architectures are not coherent, and therefore are expensive to operate, is that very little thought has gone into how components, for instance application systems and data bases, will fit together in service of the overall business. As a result, overlap has caused significant misunderstandings concerning the role of specific data bases or systems, leading to significant redundancy, inaccuracy, and expense.

John Zachman's Information Systems Architecture provides a useful perspective on information architecture. I would like to outline the 3 basic components that any architect should be concerned with. The first component is a data architecture which defines the

"things" or "what's", about which the business will store data. These things include people, places, things, concepts, and events. For a data architecture to be complete, it is quite important to include all of these types of things in the scope of a data architecture. For example, if events are left out, it is difficult to portray the major business transactions which the commercial or government enterprise may be engaged in. While it is possible to map all of the data elements represented in a transaction to other things, it is also important to recognize the value and contribution of recording the transactions themselves, since they frequently underlie the purpose of many organizations. This inclusion of events in a data architecture is particularly important for any enterprise which is involved in financial activities.

The second component which is important to information architecture is the functional architecture. This focuses on the processes performed by the enterprise, essentially the "how's". During analysis, we focus on the higher level view of the "how's", which we frequently call "what's done". This functional architecture can be impacted by changes technological advances(?) and that's something we are experiencing today at Connecticut Mutual. To say that the functional architecture of a business is really not impacted by changes in technology is simply to overlook opportunities which are presented by new technology. An airplane is more than an oxcart with wings and an engine!

The third component is the communications architecture. Essentially we represent the places of interest to the enterprise in this architecture. This is particularly important to information systems professionals today, since network computing is going to place constituent pieces of a data architecture and functional architecture at different places in the business. In fact, well designed applications in the future will distribute both data and processing appropriately. Eventually we will be able to move this processing and data around without it impacting the end user of a given application. Theoretically we can do this today, but it is extremely difficult to accomplish.

The information architecture is a model of the business itself, as well as a model of the information systems supporting the business. Why would anybody need a new information architecture? A new information architecture is needed to align a business to compete effectively in the information age, or allow a public enterprise to provide increased services to the public at a lower cost. Organizations that recognize the promise inherent in the technology available today have determined that, in fact, it is necessary to have a new information architecture. As John Rockart proposed, we should not only look at existing business processes needing automation. That's only one side of the cycle which is now speeding up more and more. The other side of the cycle is that increased technology itself is allowing and enabling, if you like,

new or changed business processes. That is to say, new businesses are actually possible and new business processes are enabled by the technology that we have today.

Philosophically this is a very different approach than I would have had 10 years ago, but when I drive up to an Avis terminal to return a rental car and I see a person standing there with something that looks like a walkie-talkie, but which is really a wireless computer terminal, it occurs to me that some fundamental shifts in what is possible have occurred. We can now construct business processes that exploit the value of time to our clients.

What we are looking for with a new information architecture is to align information systems, which I'm going to call "IS", with the business in a way which is much more dynamic than in the past. Because our task is not just to find out what the business needs or to do research in that area and then model it. It is, in fact, to discover opportunities for what the business can be doing in the environment we now have, and re-engineer the business itself. Based on my own experiences over the years with large organizations that are spending hundreds of millions of dollars to implement a new information architecture, and Connecticut Mutual which is spending considerably less, I would say that we are, in fact, re-engineering the business itself. Our objective should be clear--we need to provide new products and services to the marketplace, whether it is a public or private organization, faster. We need to do service transactions, for instance changing the address on a customer's policy when a customer moves, much faster than before. We must lower our unit costs, not by 5 or 10 percent, but 30 to 50 percent. We can do this now.

The pressure for lower unit costs will be seen in both the private and the public arena in the 1990's. To drive down costs we need more flexible businesses and more flexible systems. We are going to do need common data, some common communication networks, and some core systems which will be used across the enterprise to employ technology which is flexible and easy to change.

Some of the major enablers of this alignment between the business process and information systems includes capturing data once and having it shared across the organization. However, I can't envision being successful at Connecticut Mutual when implementing a new information architecture and aligning our business with IS unless we can provide shared data. It's a very important enabler of the new way of processing work. Not the least of which is that we want to event-driven processing in the future, not sequential processing. Event-driven or object oriented processing is going to require data which is consistent, clear and shared.

The second enabler of implementing a new information architecture is a re-engineered and automated business process. Certainly in data administration we have spent many years talking about building

business models and managing data whether it is automated or not. Philosophically I still think this is true, but it's simply not as relevant as it used to be. Consistent, automated business processes using consistent and automated data and images are where the action is going to be in the next 10 years. The ability to deliver reengineered, automated business processes is a key to justifying the use of a new information architecture.

We also need to identify what application systems are common across an organization and which application systems might be unique to each part of the organization. While it is obvious that things like personnel systems and payroll systems should be standardized across an organization, it is not quite so obvious that, for instance, accounts receivables ought to be done consistently even though accounts receivables' use of the application may be spread across an enterprise. However, I think that is exactly the situation we're looking at where we will use techniques such as information engineering to identify the activities that need to be done consistently, and then implement those activities in automation and have those activities actually performed across the organization at the appropriate time and place.

Such an implementation requires us to map the business process which is identified in information engineering with a functional view and a data view back against the workflow process at the floor level in the company. Mapping these two things together is fundamental to our future success. To do this we need an automated repository to support system design, construction, and modification. For this reason, Connecticut Mutual has decided not to wait on IBM to master a repository. We have implemented our own repository based on MSP's DATAMANAGER product. This repository links our CASE tool with our operating data bases. We are also starting to use it to hold information about our workflow at the floor level in the business.

And finally, new information technology is fundamental to the need to define a new information architecture. Object oriented user interfaces such as Presentation Manager and Windows combined with the power of multiprocessor mainframes and network communications greatly increase the options available for implementing information systems today. We at Connecticut Mutual find ourselves employing all of these tools in the applications we are currently deploying.

Let me spend another minute or two on why the new information architecture is so important. "Information Management Technology" includes large computers and personal computers. This information management technology has transformed our society and even played a major role in destroying the marxist-based societies of eastern Europe and the Soviet Union. The fundamental reason that computers have played such a major role in destroying these economies is that rational bureaucratic societies which try to plan and implement centrally can not keep up with the robust capitalist economies

because they can not create new products and services quickly enough to have their wealth enhanced. We in the western capitalistic economies, perhaps unknowingly, have moved from an industrial society based on capital and tangible goods to an information society which is largely based on symbols. For example, we don't exchange gold anymore when we buy and sell things, in fact we don't even exchange paper. What we exchange is electronic symbols that represent gold or money. It's not necessary to buy stock in a specific company anymore, you can buy a stock which represents a marketbasket of stocks from a number of companies in a specific stock exchange. These are symbols of wealth, not wealth itself, and yet the ability to buy, sell and trade upon these symbols is how service economies create wealth.

In the past, military force and money were main measures of power. We have moved now to an age where power is also based on knowledge. And if you don't think that knowledge is power, then consider a match-up between a MIG21 and an F15 which has just gone on recently in the Persian Gulf. The knowledge represented in the information systems embedded in the F15 allow that particular airplane to acquire the image of an adversary airplane while it is more than 100 miles away. Having acquired that image, the pilot then has an option whether or not to use the technology on board to destroy the adversary's airplane. The element of whether or not that other airplane can be destroyed is no longer the issue. This is simply symbolic knowledge at work.

In the financial services sector, symbolic knowledge is even more important, since knowledge of what is going on internally and externally is fundamental to success.

As we look beyond 1991 to our future as a country the use of information management technology is going to be a prerequisite to our success whether we're in a manufacturing, financial services, or government services sector of the economy.

Moving To A New Information Architecture

A number of consulting firms will probably want to talk to your organization about their expertise in helping you define your new information architecture. It is quite appropriate to use consultants, if you have not had experience in this area yourself, and you don't have people in your organization with this experience. However, defining the components of the architecture is a relatively trivial exercise compared to the process of using the new architecture. In fact, it takes so long to fully implement a new information architecture that the target will have changed by the time you get a major portion of the architecture in place, forcing you to continually revise and change the architecture as you implement it over time. With new generations of technology being deployed by vendors every three years I think it should be an assumption that the implementation of a new information

architecture is essentially a never ending, fundamental part of doing business in the future. Let me put a further definition on that issue by saying that the road to a new information architecture is what you should spend your time figuring out, since that is what you are going to be doing for the rest of your career, if you stay in this field.

Information engineering, which is a topdown approach to defining enterprise data and function, is a fundamental part of implementing a new information architecture. We simply must capture business models which are relevant and accurate, and then drive our new information systems off those. So I have a foil here which does represent a topdown perspective of moving from planning to analysis to design and into implementation. It is quite important to record what it is you are doing. If you do not know what it is you have delivered, then your ability to make changes over time will be compromised. Rebuilding the business, probably rebuilding the business continually, is fundamental to effectively competing in the information age. There are three major points I would like to make here. Information engineering for its own sake is not really the point, but information engineering to accomplish business objectives and opportunities is extremely useful.

To give you a couple of examples, one very large company, they're a household name, used information engineering to redefine their entire relationship with their distribution channels. They used information engineering very effectively to deploy new information systems to improve that relationship with their distribution channels.

Another large fortune company focused on basic product management around the world. This company was running out of product identifiers in their existing systems. Once that starts to happen results reporting and revenue projections become a lot more difficult. So this particular company used information engineering to redefine the concept of product and then to drive creation of new product number allocation off these new concepts.

A third dimension of rebuilding the business is data modeling at a very high level as a prerequisite to any significant information systems work. It is important to use information engineering or any other technique that you can get people to accept, to define a common, standard approach to managing customer, product, and revenue data. Most enterprises do not have agreement as to what is a customer, what products are and how to represent them, and what financial information needs to be recorded in all cases. This data modeling at the high level to get the data architecture in place is fundamental to the success of both information engineering and implementation of a new information architecture.

Another important component on the road to implementing a new information architecture is to have a clearly defined and

supportive information management infrastructure. This includes the skills the people need to do analysis, design, and implementation in the new environments they are going to be working in, as well as having tools available to support them. CASE tools, policies, and practices or methodologies are all important. In fact, nothing could be more of a waste of money than implementing a CASE tool without an understanding of what methodology underlies that CASE tool.

Let me put in a plug however, for perhaps the most important tool of all, and one which has not been taken very seriously by the IS community so far. A repository, or an extended data dictionary/directory, is fundamental to our success. As we move toward distributed processing, distributed data, and network computing, we can not manage change in that environment without accurate knowledge of what components are deployed at what location. In a large corporate environment with a number of systems and databases in operation, the use of a repository is fundamental to success.

The transition from the current environment to a new information architecture is very difficult and expensive. In fact I think back to three or four years ago to a large corporation which was planning to spend $300,000,000 to redefine their information architecture. People there were talking about this transition as changing the wheels on a moving bus. I think they were underestimating the difficulty of the task. My own perspective would be that what they were probably doing was rebuilding a moving bus! For that reason we need a view of the existing and future environments which will allow some insulation between the work place and the systems and databases that are in operation at any one point in time.

The following diagrams show you the framework we are using at Connecticut Mutual. There are four important components to this framework. I think most of you will recognize the bottom component as our data environment. Luckily for us, we do have a shared agents' database and a customer database that's been populated from the bottom up. The next level up is the systems or applications that are currently in place. These are changing rather dramatically--you see on the left some of the stovepipe-like CHIPS and PALLM that we implemented in the past and on the right you see our new Customer Service environment which is partially done and partially under construction. The level above that is I think one of the keys to our transition. We are using frontware to mask existing systems from the business clients who use them. A good example of this software is Easel which is currently being used rather extensively by IBM clients to mask CICS systems. This is exactly what we are doing and the implementation of a common user interface based on OS/2 or Windows masks the implementation under the covers of our existing application systems. For example, someone who is doing claims work at the event of someone's death,

might have to sign on to seven different systems to find out if an individual was insured with the company. To avoid them having to do that, this one mask takes the transaction that they enter on their intelligent workstation and drives that out to seven transactions to seven different CICS Systems and brings back the results. This is showing dramatic improvements in productivity in the Claims area, and we are now moving to implement this masking across our Customer Service areas.

The workplace which is the top level is what the people in our Customer Service, New Business and other operational areas think of as their business. At that level we see a number of different steps which occur and which move work from one place to another. At this level we are utilizing object-oriented user interfaces such as Windows as a common way of interfacing with all the Customer Service oriented systems. We are also utilizing intelligent workstations to deliver Image Processing to the workplace, and to revise the workflow itself without changing the underlaying systems.

This framework will allow us to make changes in our workflow and improve our day to day productivity without being constrained by our existing, frequently difficult to use, application systems.

Later, as we begin to implement new application systems this masking layer will help us decrease the impact on the workplace of these changes. We see what we are doing as very much reengineering a moving business to utilize share databases. As we reengineer these business processes we will increasingly see more cost effective and more complete application systems available to support the business. We will use the masking level in the transition framework I showed you earlier to insulate the workplace from changes in today's systems.

Consistent implementation of reference data bases is fundamental to our success. Reference data, data about customers, policies in-force, products, and agents are not really volatile, particularly in life insurance. People keep these life products for from 5 to 10 years, so that even policies in-force while they are consistently added to is (?) frequently and sometimes dropped, are reasonably stable. Certainly our products and our agents need to be consistently recorded and kept for some time. In fact, information about the constituent parts of our products are required to be kept by law for seven years from the time at which something is sold.

Our customer database has been loaded from the bottom up. We scrubbed and loaded this data base from existing systems, we replaced existing alpha search systems from across the company to point to these databases, and we used an architected approach to defining what the structure of these new databases should be. The concept we are trying to implement is an original, single source

for this data, obviously we can not do that until we reengineer our application systems. By the way, our product database that you see up there is at the architecture stage, it will be under construction later this year.

One of the reasons that getting reference data organized to be consistent is that we want to enable faster system deployment, not just system development. We have already seen this happen late last year when our customer database helped us put into place a customer service tracking application very quickly. This application sits on top of its own database and also the existing customer database which is running under DB2. We process around 12,000 transactions a day in this customer service tracking system.

A similar agency service tracking system is under consideration, largely because of the success of the Customer Service tracking system. Let me summarize our efforts to insulate the workplace from change as we implement a new information architecture. 1) We plan to deploy intelligent workstations (IBM PS2s) that implement the microchannel architecture. Microchannel architecture gives us plug compatibility between IBM and third party suppliers of different components of the workstation. In the future we can both upgrade and downgrade the same workstation and that is a significant benefit over the old personal computer of five years ago. 2) We will be compressing and redesigning our workflows, to increase quality, to eliminate unnecessary work and to make us easier for our customers to deal with. 3) We are going to mask existing systems with a common user access to decrease the difficulty business people have in learning how to use our systems and to insulate the workplace from changes in the underlying systems. 4) In the short term we are going to employ Image technology because we have a very paper- driven business, and we think the time is here for image technology being delivered through token ring networks to these intelligent workstations. When we finish, our business won't look like this foil, but it will look different than it does today.

What I wanted to do was to highlight a concept here of what a reengineered business might look like conceptually. The small boxes you see between the supplier and the customer are design, engineering, purchasing, manufacturing, distribution, sales, and service. Not a perspective which lends itself to success in the information age. I would propose that we will, like many of you, use information engineering to simplify and rationalize these processes and the data that supports them, so that a broader and more integrated approach is taken to our business. This particular foil summarizes these things into product development, product delivery and customer service and management. The point is that we will be reengineering the business to take advantage of new business opportunities.

To address business process reengineering we did a Business Area Analysis in 1990 in the New Business area. We will be implementing systems in that area during 1991. Also in 1991 we will be addressing receivables management.

As we succeed we will see a number of impacts upon our business of this new information architecture. Let me quickly highlight these. 1) We will be able to give service to our customers faster than before and we will be able to do that this year in many areas. 2) The quality of the service we provide will be higher, with fewer errors and much more aligned with the expectations of the persons calling in requesting service. One of the reasons the service quality will be higher is that the work at the floor level is going to be redesigned to give individuals more control from the beginning to the end of a particular transaction. As we redesign workflows, we will be achieving significantly lower unit costs in many areas. In fact, these lower unit costs have already been translated into budget targets in many of the business areas which are the subject of these systems.

Based upon my experience with other organizations, I would imagine that anyone implementing a new information architecture can anticipate some organizational realignment. We have not planned for that at Connecticut Mutual, although it has also not been ruled out. Typically when examination of the business process itself takes place, and a new architecture supporting the business is implemented, an organization may want to take advantage of what has been learned to realign the business structure.

Implementing a new information architecture is also causing a change in the relationship between information systems and the business. It is not changing because somebody wills it changed, it is changing for practical reasons. This relationship is changing from one of IS servicing the needs of the business, to IS becoming a partner with the business in changing the business itself and its underlying systems.

Information systems themselves are going to be impacted. We are going to see share databases in place, as we have already seen them be put in place at Connecticut Mutual. We are going to see the acceptance of information engineering as a common methodology for system development. We are going to see shared communication networks, if we don't already have them. And I think we are going to see decentralized development using common support services. This is already occurring at Connecticut Mutual, and I think it is increasingly true in many businesses today.

Summary

While defining the components of the new information architecture maybe relatively easy, the transition to a new architecture is not easy. We have focused on this transition because it is critical

and it is never ending. As we create models of the business we are not trying to investigate what the business needs to do, we are trying to transform what the business does. This is the essence of the partnership between IS and the business. It is not reasonable to expect business people alone to be able to understand and convey new processes which have been enabled by information technology without a considerable amount of assistance from information systems professionals. And finally, the impact of implementing a new information architecture is to provide faster, better and less expensive services and products for your enterprise. A lot of organizations in the world are pushing very hard to achieve these same results. The ones that succeed at implementing a new information architecture will have gone a long way toward ensuring their enterprise's success.

THE IMPACT OF IMPLEMENTING

A NEW

INFORMATION ARCHITECTURE

Ron Shelby
Connecticut Mutual Life

Gaithersburg May 14, 1991

CONNECTICUT MUTUAL LIFE

17th Largest	Life Insurer
Revenues	$3 Billion
Assets	$11+ Billion
Employees	4,000+
IS Professionals	
- Central	190
- Distributed	250(?)

WHAT IS INFORMATION ARCHITECTURE?

Architecture Components

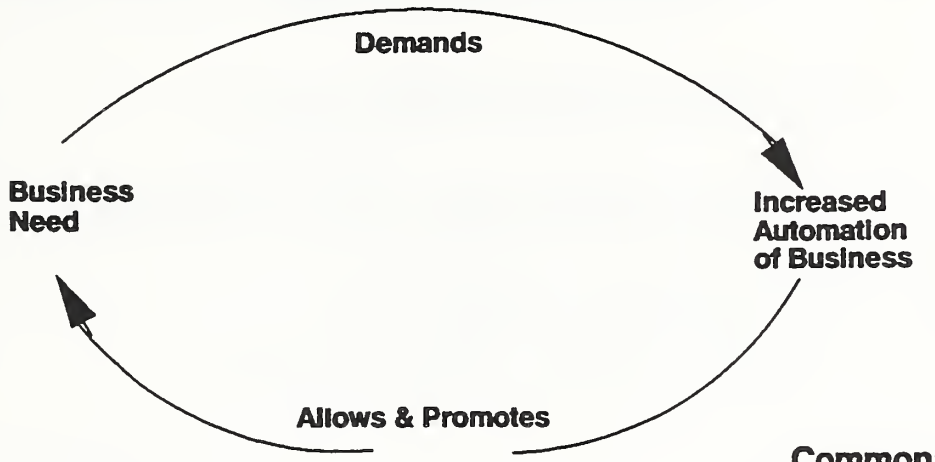
COMPONENT	DESCRIBES
DATA ARCHITECTURE	THINGS
FUNCTIONAL ARCHITECTURE	PROCESSES PERFORMED
COMMUNICATIONS ARCHITECTURE	PLACES

WHAT IS INFORMATION ARCHITECTURE?

A view of the information-related components of an enterprise which communicates the role of each component in operating the enterprise.

WHY A NEW INFORMATION ARCHITECTURE?

Align IS With Business



- ▲ Time to Market
- ▲ Faster Service
- ▲ Lower Unit Cost
- ▲ Flexibility

Common:

- ▲ Data
- ▲ Networks
- ▲ Core Systems

WHY A NEW INFORMATION ARCHITECTURE?

Enablers

- Data Captured Once and Shared
- Reengineered, Automated Business Processes
- Common Systems Used Across Organization
- Automated (Repository Supported) System Design, Construction, Modification
- New Technology

WHY A NEW INFORMATION ARCHITECTURE?

Role of Information Management Technology

- Computers Destroyed Marxism
- The Information Age = Age of Symbols
- Knowledge Is Power

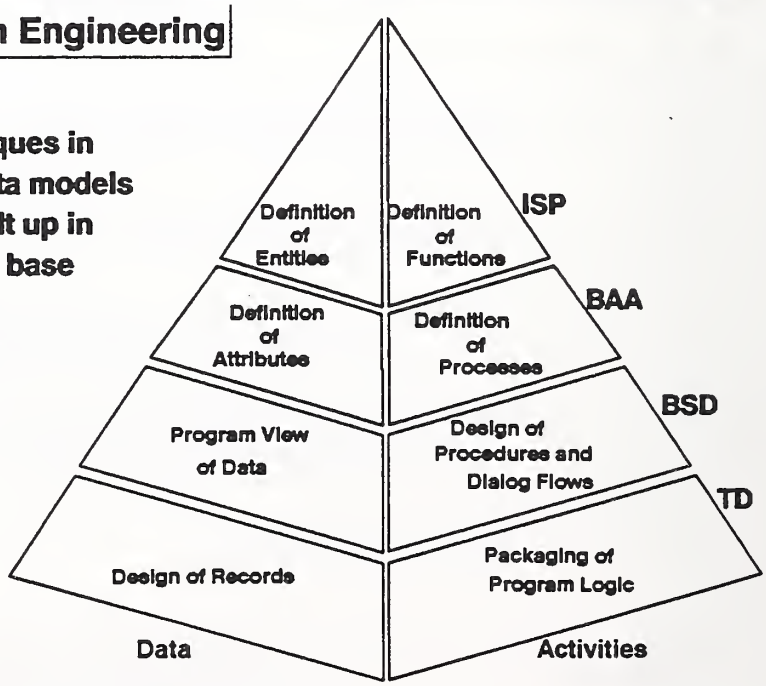
F15 vs MIG21

Using Information Management Technology effectively is a pre-requisite to success.

THE ROAD TO A NEW INFORMATION ARCHITECTURE

Top Down Information Engineering

An interlocking set of techniques in which enterprise models (data models and process models) are built up in a comprehensive knowledge base and are used to create and maintain business systems.



THE ROAD TO A NEW INFORMATION ARCHITECTURE

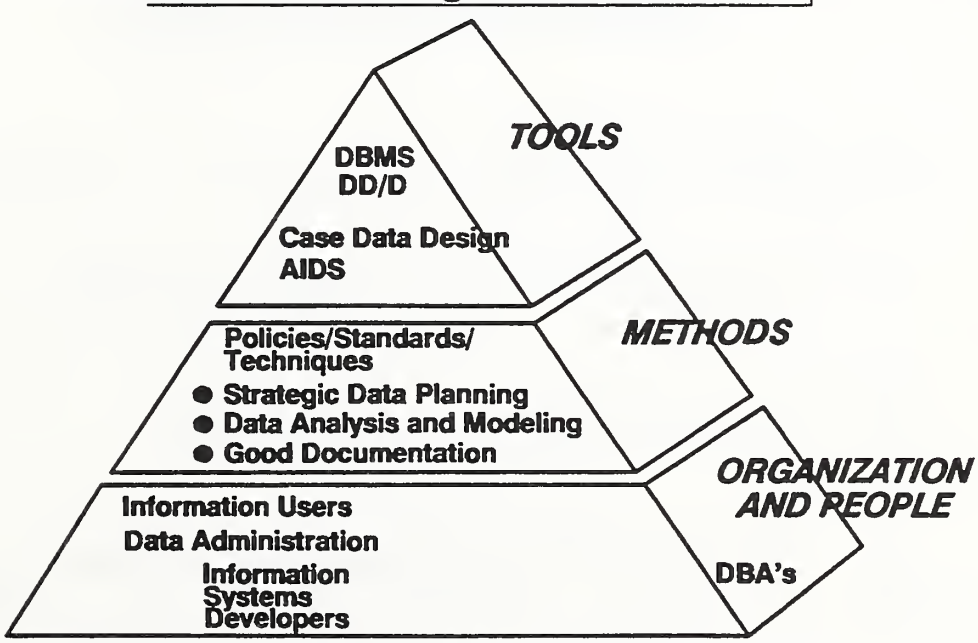
Rebuilding The Business

- BUSINESS OPPORTUNITY DRIVEN IE
- REDEFINE:
  - Distribution Channel Relationship
  - Product Concepts
  - Marketing Data Bases
  - Sales Prospecting Process
- DATA MODELING

☞ WE WERE REBUILDING SUCCESSFUL BUSINESSES

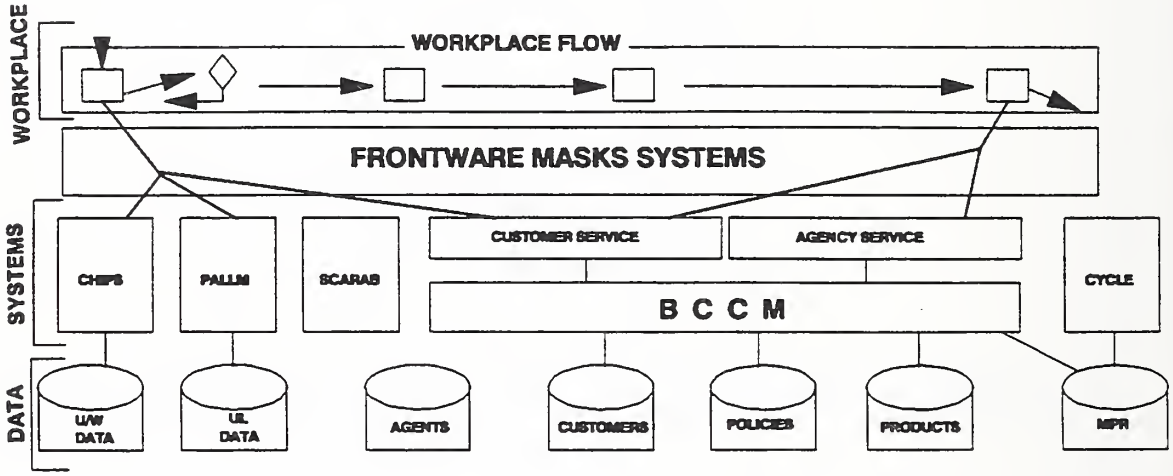
THE ROAD TO A NEW INFORMATION ARCHITECTURE

Information Management Infrastructure



THE ROAD TO A NEW INFORMATION ARCHITECTURE

A Cost-Effective Transition



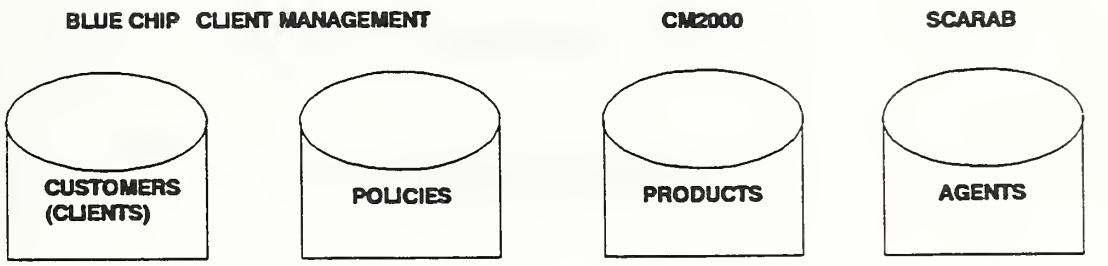
CM'S APPROACH

Reengineering A Moving Business

- Shared Data Bases
- Establish A Service-Oriented IRM Organization
- Reengineer Business Processes With IE
- Insulate The Workplace From Today's Systems

CM'S APPROACH

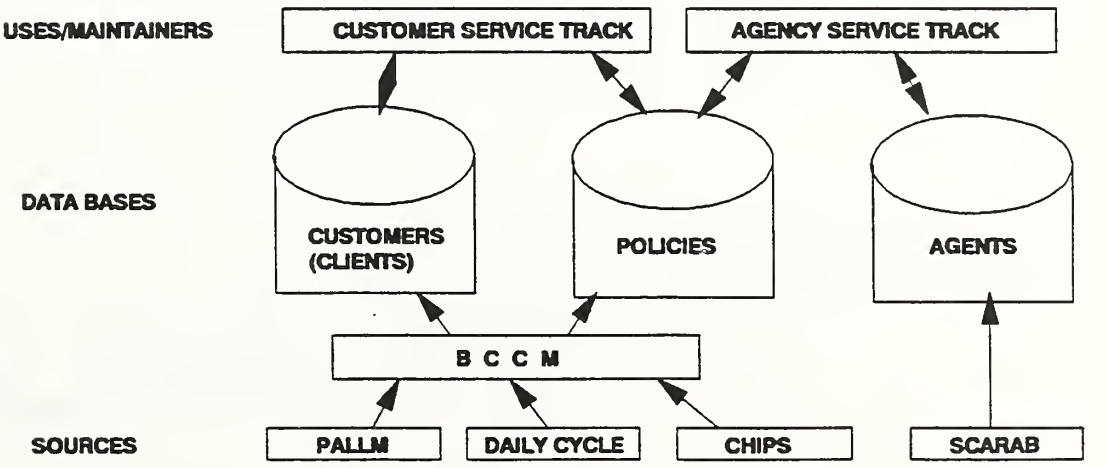
Reference Data Base



- ▲ Scrubbed and loaded from existing systems
- ▲ Replace Alpha search systems
- ▲ Architected
- ▲ Original, single source
- ▲ Multiple uses

CM'S APPROACH

Enabling Faster System Deployment



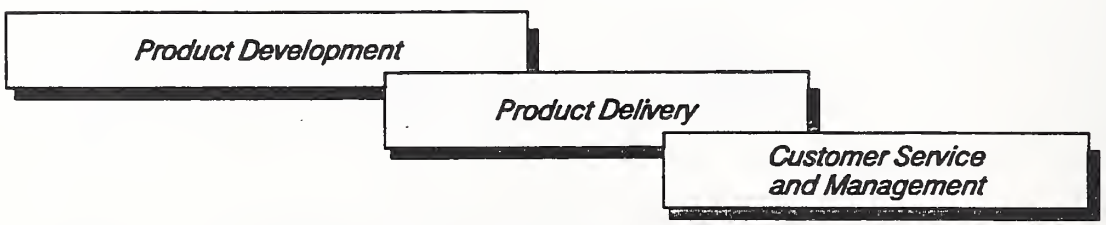
CM'S APPROACH

Insulate The Workplace From Change

- Deploy Intelligent Workstations
- Compress Workflow Steps
- Mask Existing Systems W/CUA
- Employ Image Technology

CM'S APPROACH

Reengineer Business Processes With IE



CM'S APPROACH

Business Areas Reengineered With IE

1990 New Business (Pre-Sales)

1991 Receivables (Post-Sales)

1992 Product/Customer Servicing (Post-Sales)

IMPACT OF A NEW INFORMATION ARCHITECTURE

Business Impact

- *Faster* Service
- Higher Quality
- Lower Unit Costs
- Organizational Realignment
- IS/Business Partnership

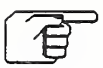
IMPACT OF A NEW INFORMATION ARCHITECTURE

Information Systems Impact

- Shared Data Bases
- Acceptance Of IE
- Shared Communications
- De-Centralized Development Using Common Support Services

IMPACT OF A NEW INFORMATION ARCHITECTURE

Summary



DEFINING IS EASY, THE TRANSITION IS NOT

DON'T INVESTIGATE THE BUSINESS, TRANSFORM IT!

FASTER, BETTER, LESS EXPENSIVE SERVICES MUST BE A DELIVERABLE

Fred L. Forman

Topic: Data Planning - What Good is it Any How?

Description: Frequently, data planning becomes an end unto itself. Data modelers sometimes spend years without adding business value to their organization. But somewhere, sometime all that could change with little or no warning. Be prepared!

Biography: Fred L. Forman is American Management Systems' (AMS) chief technology officer and manager of the Corporate Technology Group, a group of senior technical specialists responsible for providing specialized technical support for large client projects, for the overall technical quality of AMS client and internal development projects, and for selecting and developing systems planning, development, and project management methods. In addition, he is a member of AMS Management Policy Committee.

Dr. Forman has over 20 years' experience in preparing strategic systems plans, developing large-scale information systems, designing telecommunications networks, evaluating system software and hardware products, determining system capacity requirements, and managing large computer centers.

Dr. Forman has been with AMS since 1971. He hold a Ph.D. in physics from the University of Pennsylvania. He has published numerous articles and frequently lectures on systems development productivity and technology futures. He is also an active member of several professional societies.

Once upon a time there were two far away countries - Dataland and Realitania. The people of Dataland, known as Datalandians, were ruled by the tyrannical Data Administrators, although most Datalandians were very friendly mild-mannered individuals who were content to live their lives refining their subject data bases, enterprise models, and entity-relationship diagrams.

The heathens of Realitania, on the other hand, were impatient, indecisive, and generally ill-tempered. They were ruled by individuals called The Management. It will come as no surprise to you that the Realitanians were largely interested in acquiring new lands as a way of expanding their evil empire.

The story I want to tell you about is what happened when the heathens of Realitania invade Dataland, a struggle that was over in a very short time, the data dictionaries of the Datalandians proving no match for the application packages of the Realitanians. But I digress. Let me first go back to long before the Great Data Architecture War as it came to be known, to a time of peace and joy in Dataland.

DATALAND

Dataland was a place where all idealisms came true. A land of massless strings, frictionless pulleys, perpetual motion machines, unlimited budgets, and cooperative users. In short, it was heaven on earth. The primary currency of Dataland was information, and companies with the most information were, quite naturally, the most successful and highly regarded. Little wonder then that the need for strategic data planning was unquestioned - for after all, everyone knows that data is the heart of all information. Indeed, data was so highly valued that there was no more serious crime in all of Dataland than data rustling, not even the failure to put all data records in third-normal form.

Naturally in such a land of perfection, what could be more important to one's career than obtaining perfection in one's data models - perfection in both form and method. Lowly information workers could only marvel at the wondrous and intricate diagrammatic methods that their seniors could invent to describe data in yet ever more detail. Their apprenticeship consisted of many years in trying to use their tools - data dictionaries and DBMS's - to make these elegant creations work.

But these primitive tools were no match for the genius of the best of the data planners, the Data Administrators. In a never ending game of king-of-the-hill, information workers and Junior Data Administrators tried to create new methods and forms to humble the Senior Data Administrators.

I've brought a few examples of the high art of Dataland with me tonight. This first example (Exhibit 1) won the gold medal in the

1986 Data Design Olympics in the "creative use of shading" category. For example, notice the subtlety in the relationships between Skill, Project, and Employee.

This next example (Exhibit 2) won the Royal Order of the Datum award for 1985. This diagram became a legend in its own time throughout all of Dataland because it was prepared by a Junior Data Administrator. Yet, this one diagram has ensured a place in the Data Administrator's Hall of Fame for its creator. Data modeling critics have compared the author's creativity to no less notables as Beethoven and Einstein.

My third example (Exhibit 3) looks ordinary at first, but if you look more closely, you'll begin to appreciate why this beauty made the cover of Entity magazine. Notice, for example, that the "circle-s's" only go in, never out; notice the variation in the thickness of the lines; notice the two arrows between Sales-Force and Leads. I could go on, but I hope you can begin to appreciate just what a masterwork this is.

Of course, you will not be surprised to learn that in a land ruled by Data Architects that there was no place whatsoever for application packages.

It was said that there once was a renegade band of system developers that tried to develop a general purpose payroll system, but that the Ruling Council of Data Architects immediately saw what a threat such an idea was to their world of perfect form and methods, and they banded together to ensure that technical evaluation criteria were always set to exclude the consideration of such application software packages in the first round of any competitive procurement. To do so was really quite a simple matter. All that was necessary was to make the most important selection criteria based on conformance to local naming conventions, module size restrictions, system usage conventions, and the like. If that didn't stop the threat, then they could always insist that the system documentation fully comply with local standards, and you've just seen some examples of what that means.

In any event, it has been many years since the threat of general purpose application packaged software was significant.

Another triumph of the Data Architects was that new systems were only redone all at once in a single massive development project. Furthermore, it was a convention in Dataland that these new systems would all be turned on at once, on October 1, which was celebrated throughout the land as Start-up Day, to coincide with the Datalandian federal fiscal year. No one seemed to remember exactly how it had come to pass that the federal fiscal year began on October 1; it was just one of those aberrations that everyone had to accept.

In addition, data conversion from old systems was against both the law and social tradition. Hence, the Data Architects could be sure that new systems would fully and faithfully implement their integrated data architectures.

You may be wondering what happened to that class of people commonly called the Users in Dataland. Alas, they were of a lower social order, relegated to the relatively menial task of maintaining the data in the integrated data bases. Only they had to deal with the small imperfections of life, such as payments with no invoice number. Alas, the poor Users were never to rise to the important ranks of the Data Architects.

REALITANIA

Now let us turn our attention to life before the War in the land of Realitania. In many ways it was the antithesis of life in Dataland.

As I mentioned before, enterprises in Realitania were controlled by The Management, individuals who were rarely actually seen, communicating their directives primarily via memos, although sometimes one's supervisor would say he talked to Management and was relaying a directive verbally.

Of course, as you must already know, the driving force for all decisions was money. Or rather, to be more precise, saving money in overhead operations, such as data processing. Incremental changes were only made when they were cost-justified, regardless how intuitively sensible an idea might be. Those who could not cost-justify their contributions to the organization were summarily dismissed from their present jobs, often ending up doing applications maintenance. Indeed, it was sometimes said that over three-fourths of all applications staff did nothing but systems maintenance, a truly frightening statistic, if true.

In most organizations the status of the data bases could best be described as fragmented, dispersed, unintegrated, and undocumented. This was not surprising given that many key operational systems were typically 10, 15, or even sometimes 20 years old. Somehow it seemed that the analysis always showed it was more cost-effective to keep doing maintenance than to do a new system. Not a surprising result when systems support was viewed as a bunch of two, five, and 10 day projects.

It could be said of Management that at least they were consistent in their ideals. Indeed, many of them drove unusually old cars because it was almost always true that the cost of the next repair was significantly less than the cost of a new car.

Of course, in Realitania application packages abounded. There were seemingly hundreds or even thousands available to do almost any

function -- payroll, accounting, loan processing, billing, accounts receivable, general ledger, income tax, and on and on. Indeed, these packages were often targeted at very narrow niches in the marketplace, such as a fixed assets system for steel companies in bankruptcy.

It will not surprise you that the decision as to whether or not to use an application package in an organization was usually imposed from The Management when they learned that the application package would cost $100,000 and custom development would cost $500,000. No matter that the package used ISAM while the rest of the corporate systems used IMS. And no matter that the changes in procedures to fit package limitations and restrictions would ensure weeks of operational disruption.

Furthermore, The Management would insist that a new system be made operational on June 23 because it was the chairman's birthday, necessitating a crash $750,000 special conversion effort. It was all accepted as business as usual.

Strangely, The Management was not particularly happy with the status quo. While it was true that paychecks got out on time, and the monthly accounting reports showed up on time, nevertheless, when the Management wanted to look at profitability by product or at all relationships with particular client, such information simply could not be produced because the disparate data bases could not be brought together for such purposes.

So there was great sense of unease over this situation. In the distant past, the systems people got The Management, against its better judgment, to try some strategic systems planning, which the systems people called BSP. But after some time there was no useful result of such efforts, merely some reports that The Management couldn't understand and requests from the systems people for increasing amounts of overhead funding to continue to do yet more planning.

Now The Management saw the trap they had fallen into and quickly cancelled all such planning projects throughout all of Realitania. In fact, the phrase "strategic planning" was not socially acceptable, and it was no longer heard in the corridors of Realitanian enterprises.

The Great Data Architecture War

There really isn't much to say about the War itself. It was over in what seemed like an instant. Battlefield management programs written in 4GL's running against pure relational data bases meeting all of E.F. Codd's 12 rules (including referential integrity) simply did not provide Datalandian commanders with the timely information they needed to react to the invading forces of Realitania.

The population of Dataland was quite fearful, especially the Data Architects, because they knew of the ruthlessness of The Management and especially of The Management's disdain for all that they stood for.

It was at this point that something most unexpected happened. As is normal in a takeover situation, the new managers, in this case, The Management, insisted on going over the important documents for the Datalandian corporations and for the Datalandian government. But of course, Datalandia being what it was, the Datalandians delivered to The Management the very finest of their data model diagrams, some examples of which we have previously looked at.

When these data models were seen by The Management, they were confounded by them, much as hieroglyphics must have confounded early explorers of the Nile Valley. The Management demanded to know how to interpret these critical documents, but were only further perplexed over many-to-many relationships, third-, fourth-, and fifth-normal forms, subclasses, and the like.

On the other hand, The Management did sense that there was something important underlying all this, something that might enable them to get less data and more information. But as you can easily imagine, The Management was most skeptical because of their past bad experiences with anything that smacked of planning.

So a special Tribunal was established, and the Datalandians selected a group of their finest Data Architects to represent them. The Management agreed to give the Data Architects a chance to explain how their methods and tools could be beneficial, in a cost-justified manner, to their organizations. If the Data Architects were successful, the Datalandians would be permitted to work mostly on new applications. However, if they were not, they would have to do systems maintenance. Even worse, some might have to join the ranks of the Users.

The Special Tribunal

The Data Architects huddled to prepare their case, and it took only a few moments for them to realize they had a very serious problem. In a word, they hadn't the foggiest notion how to make such a presentation to Management in terms that Management could identify with. They didn't even really understand the language of The Management. In short, they were in BIG trouble.

One Data Architect said "let's prepare a master legend of all of our symbols and train The Management in how to read our data model diagrams. Then we can hold walkthroughs, and they will come to appreciate the importance of our works." But, fortunately for them, one of the Data Architects had once visited Realitania. He had tried this tactic on The Management, and it was a total failure.

The Data Architects had been given two days to prepare their case, and they pretty much frittered away the first day with one fruitless scheme after another. By the end of the day it was clear that, no matter how well intended, approaches which depended on making The Management more like Data Architects were not going to succeed.

Late that first day one of the Data Architects made a fateful suggestion: "why don't we talk to someone from The Management before we present our case." The Data Architects approached the Tribunal with this request. At first the Tribunal was reluctant because they knew that The Management were very important people and simply did not have the time to spend on such matters as planning. However, given the special circumstances, they did appoint one member of The Management to spend the next day in helping the Data Architects prepare their defense. His name was Bob.

So bright and early the next morning Bob met with the representatives of the Data Architects, and, fortunately for them, he was a person of great heart who truly wanted to help these hapless souls. Quite naturally, Bob's first question was to ask the Data Architects just what they did do! After 30 minutes on normalization, SQL-operations, many-to-many relationships, and the drawbacks of hierarchical data bases, Bob realized he was getting nowhere.

So he decided to take a whole new tactic. "Let's try and take this from the top-down," he said. "I'll tell you the kinds of things that ring The Management's chimes, and let's see if you can help with any of them." All agreed this was perfectly logical way to approach the task at hand.

The Language of Management

Bob said: "it's really very simple, in principle. It's just that in practice things get pretty complicated. Corporate management wants to improve overall profitability, both on an absolute and on a relative basis. In governmental and non-profit organizations there is a similar principle operating, even though those types of enterprises do not directly generate profits. Rather, they strive to improve overall efficiency, or the ability to reduce the unit costs of services. In effect, management strives to sell more while spending less."

At this point, one of the Data Architects interrupted Bob and said: "that just sounds like buy low, sell high; I don't see how that helps us defend the benefits of data planning. Anyhow, I think this is all crazy. It's obvious that integrated data is better than fragmented data."

"Not to The Management it's not," said Bob. "Over many years they have spent millions and millions for integrated data, and all they've got to show for it is shelfware. So maybe the theory of integrated data is better than the theory of non-integrated data, but in reality, getting integrated seems to have eluded all the enterprises in Realitania."

"Anyhow, let me get to the level at which we can try to identify where you might be able to help," said Bob. "One way to reduce costs is to find ways to do current jobs more cheaply, such as by reducing the cost to produce each paycheck. Reducing internal unit costs has been the primary area to which data processing has addressed itself in the past, building systems to support internal operations, such as accounting, sales, manufacturing, and so forth. Do you see any other ways that you can further reduce internal operating costs?"

"Yeah," said one of the Data Administrators, momentarily forgetting the gravity of the situation, "lower the salaries of the Users!" Well, as you can imagine, that comment provoked quite an uproar, and it took several minutes to get the meeting back on track.

Unfortunately, the general consensus among the Data Administrators was that there wasn't much they could directly influence via their planning to improve the efficiency of internal operations, since that seemed to imply working with existing systems, which, as I mentioned before, was strictly forbidden in Dataland. Of course, that type of thinking was quickly squelched by Bob, who reminded them that they were now operating under the laws of Realitania. Furthermore, if The Management behaved as they had in prior conquests, they would make Datalandian enterprises use the same systems as were currently in use at Realitanian enterprises -- existing old systems, application package, and all.

This was quite a jolt to the Data Architects, who had not up to that point really internalized the difficulty of their situation. "Could you give us some examples of the types of data problems you have with current systems?" asked one of the Data Architects.

"Well, in my company, a financial services company, one of our major problems is in pulling together all of our relationships with each customer. A particular customer may have a car loan with us, a checking account, a certificate of deposit, and more. Yet because each of these "products" is handled by a separate system that has been installed one-at-a-time over the past 10 years or so, there is not a single unified customer number across all systems to help us pull the information together. Today, we have unsynchronized, often inconsistent, information. For example, when we get an address change, we have to enter it five or ten times into different systems, and even then only when the customer explicitly tells us for each system. Can data planning help with that situation?"

One of the Data Architects said: "we don't have those types of problems here in Dataland because we just start from an integrated relational data base and all new systems, so we're not experienced with that type of problem."

But a more thoughtful Data Architect said: "Couldn't you create a new customer file, call it the Consolidated Customer File, based on integrated, normalized data design principles, where all the customer information was consolidated with references back to the existing systems?" He then sketched out what he had in mind (see Exhibit 4).

"But how will the Consolidated Customer File get built and integrated with the other systems?", queried Bob.

"Well, to start, you could take all the customer records, across all the systems, and build a matching program to look for exact duplicates -- that is, where first name, last name, street address, city, state, zip code, and home telephone were identical -- and initially build the consolidated file based on that. Then you could sort the lists of remaining customer records by various keys and manually look for other matches. Over time, each system could be modified by putting in a standard Customer Number lookup routine in place of direct calls to the system-specific customer file. This routine would first go to the new Consolidated Customer File to get the information."

"Obviously, there's a lot more work needed to figure this out in detail, but it seems like it would work. Then, over time, you would be building a Consolidated Customer File, even if you were slow to install new systems. And when it came time to install new systems, whether or not based on an application package, you could make sure that they use the Consolidated Customer File from the outset."

Bob was impressed. "That's the ticket," he said, already envisioning hundreds of other benefits, some even quantifiable, from having a single integrated customer file -- new sales leads, reduced mailing costs, lowered credit risks, reduced maintenance costs, and many more.

Now Bob turned to another topic. "A second way to improve profitability is to reduce the unit cost of supplies," said Bob, "do you have any way to help us in that area?"

The same Data Architect that said he didn't know anything about existing systems replied: "how can we affect supply costs? That's for the manufacturing people to negotiate with the suppliers, not for us to deal with."

But he was interrupted by another, more thoughtful, Data Architect: "I once read about a firm that was able to reduce the ordering and

delivery times for supplies by creating direct data links with their main suppliers. As a result, their company was able to keep a much lower level of supplies on hand, thereby reducing their warehousing and inventory costs."

Said the first Data Architect: "But is it our job to make things like that happen?"

Said the thoughtful one: "First of all, why isn't it always our responsibility to find more effective ways of using our data. After all, haven't we been preaching that data is a corporate resource for many years?

What other interpretation could we put on our responsibility than to find more effective ways to use that data. Unfortunately, we've overly focused on the formal design part of our job and underfocused on our overall enterprise-wide mission."

"Second, regardless of who is eventually put in charge of the project to build such systems, data exchange will be the key and that is unambiguously our job."

"That's the ticket," said Bob again. "But now let's turn to the most challenging area of profitability improvements, increasing sales. One of the great difficulties in the financial services business is in introducing new products, which we must continually do to adapt to changing market conditions and to offset that our competitors do. However, if we want to offer a new product, like a 1/2% break on home equity loans for customers that maintain a minimum balance in their checking account and have a safe deposit box with us, that is very difficult because of the kind of system fragmentation I told you about before. Also, we are very hindered in certain kinds of financial or marketing analysis due to this type of data fragmentation."

This time, the earlier first-to-answer Data Architect was somewhat more thoughtful: "This is the area where we can be helpful by implementing the integrated data bases that we have been defining all these years. Those are the kinds of problems that are more easily handled by creating a data base that reflects the natural business data organization."

"That sounds good," said Bob, "but how can all this help us now -- we don't have the luxury of installing all new systems every few years.

"I've got an idea," said the first-to-answer Data Architect. "Let's implement the target integrated data base in its own separate environment, using user-friendly access tools so that users can directly acquire the information they want without having to go to the systems people every time. We could build load programs to populate the target data base from the production files

and run them periodically. We could call it a Center for Information or, better still, an Information Center."

"But you won't have some of the information to fill the target data base, plus timeliness will be an issue for some kinds of uses," said Bob.

Replied the Data Architect: "True, but this would be a start. Something is better than nothing, as long as everyone is aware of the limitations. And it would clearly be more than adequate for many types of analysis and possibly even some operational uses. For example, a customer inquiry could determine eligibility for new products, even if it were not real-time data."

Bob said, "that's the ticket. I think we're ready to prepare our case now."

Making the Case

So the day of the trial arrived and, as you will not be surprised to learn, there was great nervousness throughout all of Dataland. The trial was televised, and there was not a person in all of Dataland over 15 years old that was not watching.

At exactly 10 o'clock the trial was called to order, and the person the Data Architects had selected to represent them rose to address the Tribunal. A silence came over the room; in fact the silence was all over the entire country of Dataland.

"Ladies and gentlemen of the Tribunal, citizens of Dataland, and citizens of Realitania. The Management of Realitania have given us this opportunity to explain how we can be helpful to them in the conduct of their business."

"We can be invaluable to improving organizational productivity and efficiency. We can identify and help to develop ways to use data more effectively within existing systems, such as by establishing a single consolidated client file for a financial services organization. We can work with suppliers to identify ways to reduce order time by more effective and timely information exchanges, thereby reducing inventories and carrying costs. And lastly, we can support the development of systems which are more responsive to management's desires to identify new products and to improve customer service. In short, we can be an important player in reducing costs and increasing sales."

"But there are two very critical aspects of our ability to be effective that I want to mention. The first is that we must work closely with The Management to ensure that we both help The Management identify the highest payoff projects and yet be responsive to meeting their priorities. For example (see Exhibit 5), we can help rank projects by development cost, risk, and

benefit to ensure that the organizational priorities are consistent with the most efficient use of its information resources."

Indeed, we need to really think of ourselves not as Data Architects, but as Information Architects, helping each organization to most effectively deploy its information resources. Since we will have to work closely with The Management to be effective, we recommend the establishment of a new position in every enterprise, the Chief Information Officer, who would be the overall coordinator between The Management and the information processing parts of the organization."

"Secondly, we need the realization that our goal is not perfection, but to implement practical, helpful solutions to day-to-day inefficiencies, to support the development of new systems, and to help management. While we need to have perfection as a long-term goal, we may never quite reach it, although we can be very successful nonetheless."

"So that's our proposal. Thank you very much for the opportunity to present our case."

Well, I don't have to bore you with the details of what happened next. The members of the Tribunal were unanimous in agreeing that they wanted to work with the Data Architects and take another shot at solving their short-term and long-term management problems with the benefit of more effective information systems. There was great joy throughout all of Dataland, but everyone knew their work was only just beginning.

Said Bob, "that's the ticket."

Exhibit 1

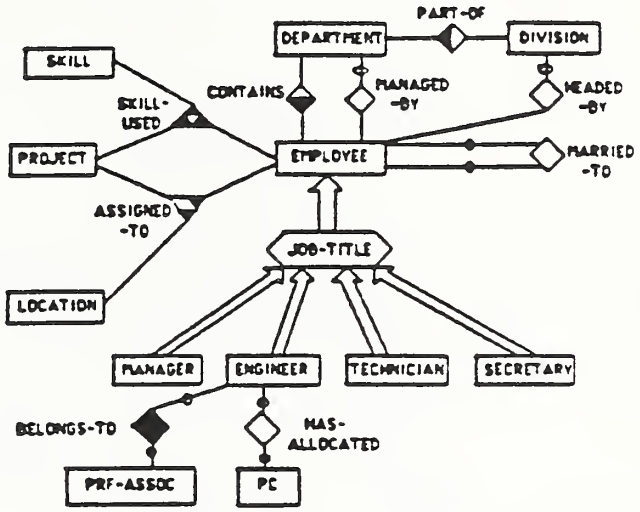


Exhibit 2

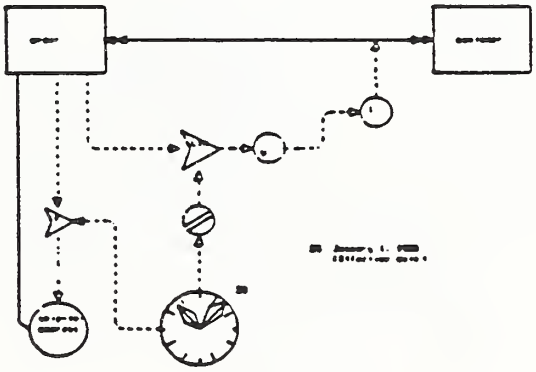


Exhibit 3

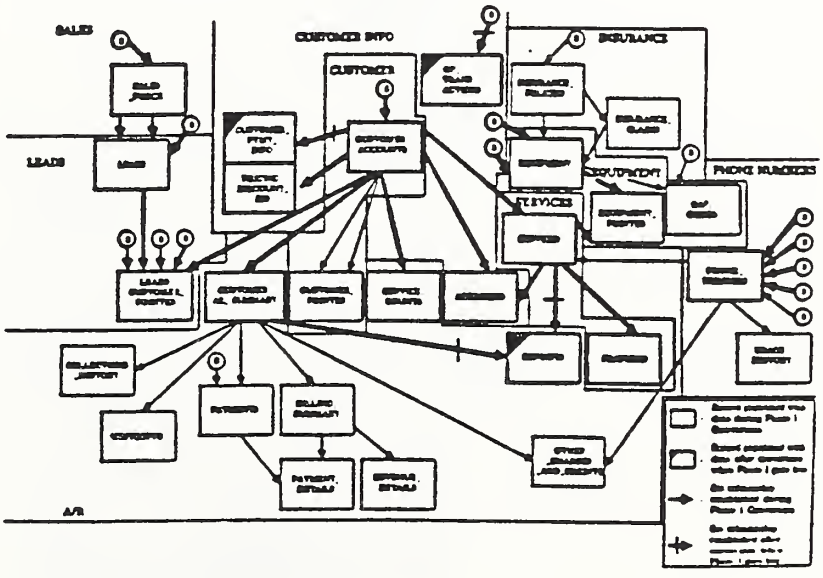
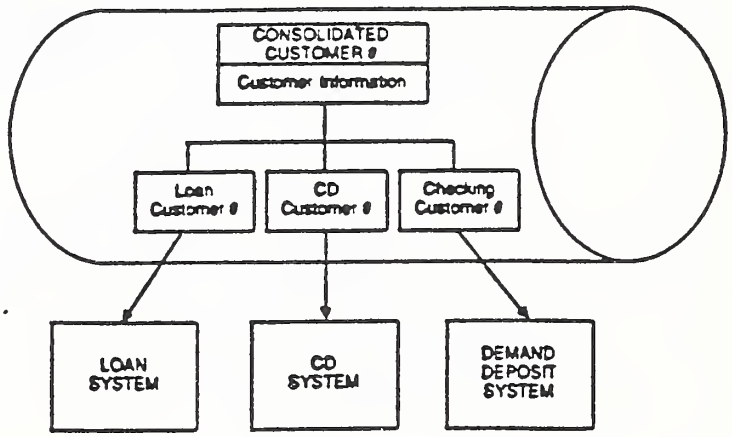
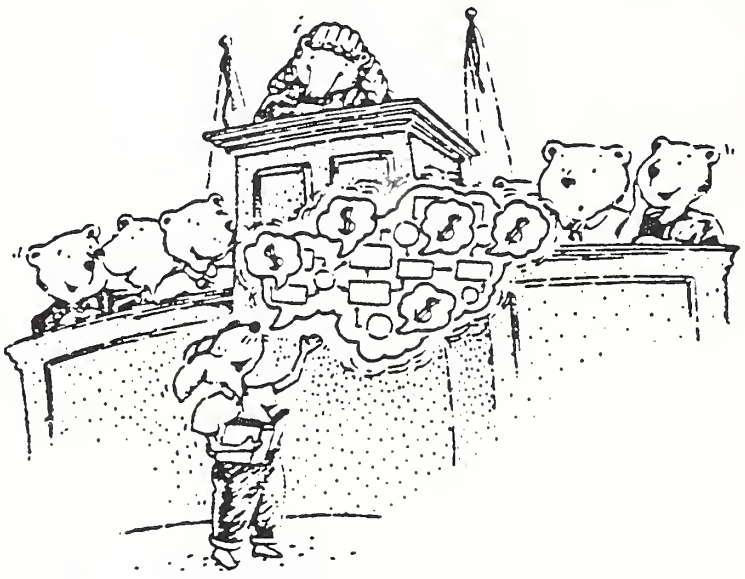


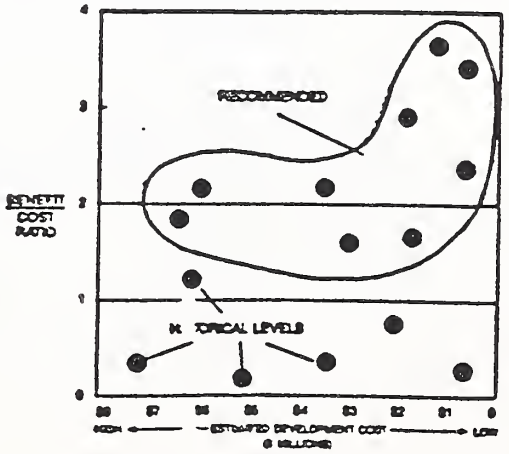
Exhibit 4





DETERMINING SYSTEM PRIORITIES

Exhibit 5



Joseph H. Oates

Topic: Data Modeling: The Data Administrator Perspective vs. the Traditional Application Development Point of View

Description: A data sharing environment reveals many issues that have not been a concern to application developers in the past. Data Administrators follow a much more rigorous set of rules in developing a global data model than is required for data models for systems that do not share data. This presentation will highlight the different perspectives and will give examples of some of the issues.

Biography: Mr. Oates has more than 20 years experience in analysis, design, programming, modeling and consulting for large systems. He has developed business systems for General Motors, Burroughs and Comshare and factory automation systems for the J.B. Webb Company. He designed a successful PC-based CASE tool. Mr. Oates has been a consultant for the past five years, including a stint with a Big Eight accounting firm. He is currently a senior consultant for Life Cycle Technology where he has been the lead data modeler on large financial management systems as well as offering courses in basic and advanced data modeling.

Data administrators have a different philosophy and perspective on data modeling than do most traditional application developers. This statement is based on several years of observation of system development projects. This presentation identifies some of the major differences between data administrators and most system developers.

Most data processing professionals have the same basic goals for systems development activities. Among them are:

- Quality Systems
- Rapid Development
- Sharing Data Among Applications
- Easy to Add Functionality
- Inexpensive to Modify
- Low Operational Lifetime Costs
- Allow Organization to Respond to Market and Competitive Changes

Incidently, the organizations who pay for systems development also have these same goals.

Unfortunately, however, the systems that are actually developed exhibit different characteristics. The characteristics of real systems include:

- Users Are Seldom Satisfied with Systems
- Data Sharing Is Still Only a Concept for Most Organizations
- Projects Typically Exceed Schedule and Cost Estimates
- It Is Difficult to Add New Functionality
- Systems Are Often Prohibitively Expensive to Modify
- Consequently, They Exhibit Very High Operational Lifetime Costs
- Too Often, These Characteristics Actually Prevent Timely Response to Market and Competitive Changes

Several studies have been done regarding system quality, schedule and cost. These studies show that the above undesirable "realworld" characteristics are true of most systems. They also show that most of the undesirable characteristics are due to errors

in the analysis stage of the software development life cycle (SDLC) such as incorrect "facts," omission of critical facts, inconsistencies and ambiguities. These studies also show that approximately 80% of the cost of maintaining a system over its operational life results from these same errors.

It is important that system developers understand the purpose, characteristics and critical roles in the analysis stage. Following is a definition of the analysis stage from this perspective:

ANALYSIS STAGE

PURPOSE:

- To Identify and Define the Objects Whose Persistence Transcends Time
  - Entity Types (This Is, After All, What It's All About)
- To Identify and Define the Pertinent Attributes for Each Entity Instance
- To Identify and Document the Semantic Associations Among Entities
  - Relationships (Existence and Definition Completeness)
- To Discover and Clearly Document the Underlying Entity Behavior (Policies, Procedures and Rules)
  - External and Internal Events
  - Entity States
  - Entity State Allowable Operations
  - Legal State Transitions
  - Relationship Rules
- To Convey These Essential Requirements to the Design Stage in Some Generally Accepted Format
  - ERD
  - STD
  - DFD
  - Other

CHARACTERISTICS:

- Assumes Perfect Technology Within System Context
- Independent of Real-World Technology
- What, Not How

- Logical Requirements View
- Data, Behavioral and Function Submodels Developed

CRITICAL ROLES

- Domain Expert
- Data Administration
- Analysts

Some authorities refer to the Analysis State as the Software Requirements Stage.

Implicit in this definition of the Analysis Stage is the assumption that data sharing is a goal shared by developers.

It has been my observation that application developers have a different perspective about systems than data administrators. The following table illustrates the main differences:

Traditional Application Developer	Data Administrator
Single System Oriented	Enterprise Oriented
Function Oriented	Data Oriented
File Oriented	Data Sharing Oriented
Physically Oriented	Logically Oriented
Algorithms	Business Rules
Access Concerns	Semantic Concerns
Presentation Concerns	Integrity Concerns

The presentation contains three examples that come from actual projects. These examples illustrate several data modeling errors that have caused specific systems to fall short of the goals for system development. There is also a listing of the problems that were caused by the way that the data was modeled.

The presentation also contains an alternative to each of the real-world examples reflecting a data administration perspective. This is followed by a listing of the benefits of modeling from the data administration perspective. The names and forms of the entities have been modified to protect the privacy of the organizations who developed the systems.

Finally, there are some observations and conclusions. Among them are:

- Many Organizations Skip the Analysis Stage and Start at the Design Stage

- Most Analysis Stage Errors Are Attributable to Data Modeling or the Lack Thereof

- The Specific Data Modeling Errors that Cause Most of the Errors Are:
  - Failure to Identify All Entities
  - Failure to Correctly Identify and Model the Entity Life Cycles
  - Failure to Correctly Identify and Document Entity State Allowable Operations as Part of the Data Model
  - Failure to Correctly Identify and Document Entity State Transition Rules as part of the Data Model
  - Different Levels of Aggregation of Data in a Single Entity
  - Redundant Attributes
  - Multiple Definitions of Entities and Attributes
  - Lack of Data Administration Participation During the Analysis Stage

Many of the same studies cited above indicate that the most effective way to identify and correct analysis stage errors is by implementing an effective review process. However, the people doing the review must know an error when they see one. It has been my experience that most developers and many data administration personnel have not had the experiences that develop the ability to tell a data modeling error when they see one. Therefore, there must be a set of quality metrics that are available and used during the review of data models and other Analysis Stage products. For example, LCT has identified more than 250 quality metrics for the SDLC, including more than 100 for the analysis stage. It was through the use of these quality metrics that the errors in the examples were identified and corrected.

There are several conclusions that I hope everyone leaves the presentation with. Among them are:

- Traditional Developers Want to Produce Quality Products

- Traditional Developers Don't Commit Modeling Errors Intentionally

- Traditional Developers Don't Have Any Criteria to Judge Quality of Products They Produce

- Most Data Modeling Courses and Books Do Not Give the Enough Detailed Quality Criteria to Be Useful to Developers

- CASE Tools Can't (Yet) Help With the Real-World Situations Just Discussed

- Worse Yet, There Is a Perception that Using a CASE Tool Makes Anyone an Instant Expert Data Modeler

- Data Architectures Can't Be Effective Unless Component Data Models Correctly Reflect Business Rules
- All Data Modeling Should Be Done From the Data Administration Point of View
- More Training Alone Is NOT Sufficient
  - Training Generally Conveys Concepts Only
- Developers Need Comprehensive Work Plans For Each SDLC Stage
- Better Developer Work Plan Characteristics Include
  - More Detailed Work Breakdown Structure (WBS)
  - WBS Based On Interim Deliverable List for Each Subtask
  - Quality Metrics on which to Judge Each Interim Deliverable
- Data Models Must Include Entity Life Cycle In Addition to Static Data Structure
- Data Administration Must Be More Active in System Development Projects
- Rather than Starting from a Functional Decomposition Point of View, If the Following Are Available and Correct, the Process Model Just "Falls Out"
  - Static Entity Model
  - Entity Life Cycle Model
  - Entity State Allowable Operations
  - Entity State Transition Rules

DATA MODELING:

THE DATA ADMINISTRATION PERSPECTIVE
VS.
THE TRADITIONAL APPLICATION DEVELOPER
POINT OF VIEW

JOE OATES

LIFE CYCLE TECHNOLOGY, INC.

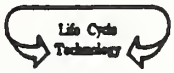
3603 CHAIN BRIDGE ROAD

FAIRFAX, VA 22030

(703) 359-9700

PREMISE

DATA ADMINISTRATORS HAVE A DIFFERENT VIEW
OF DATA MODELING THAN
TRADITIONAL APPLICATION DEVELOPERS

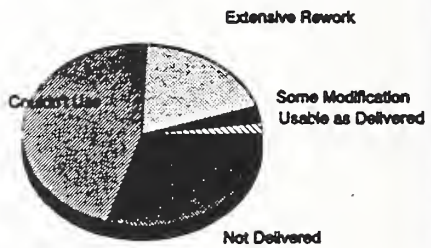


GAO STUDY RESULTS

RESULTS

- 2% Usable as delivered
  - Well understood utility
- 3% Some modification
  - Simple application
- 19% required extensive rework
- 46% Delivered, couldn't use
- 30% Not delivered

GAO ANALYSIS
OF APPLICATION DEVELOPMENT PROJECTS



Life Cycle Technology

SO WHAT?

- Critics Say
  - Study Flawed, Too Few Samples
  - It's Several Years Old, Things Are Better Now
  - We Have New Methodologies Now
  - We Have CASE Tools Now
  - We Do Data Modeling Now

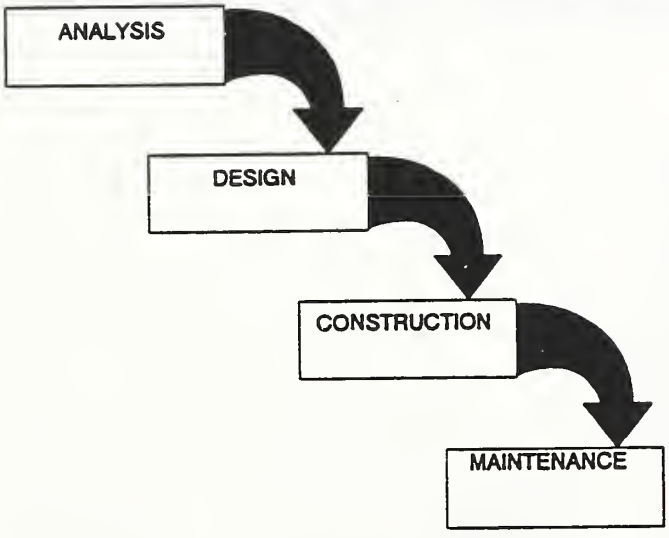
Life Cycle Technology

HERE'S SO WHAT!

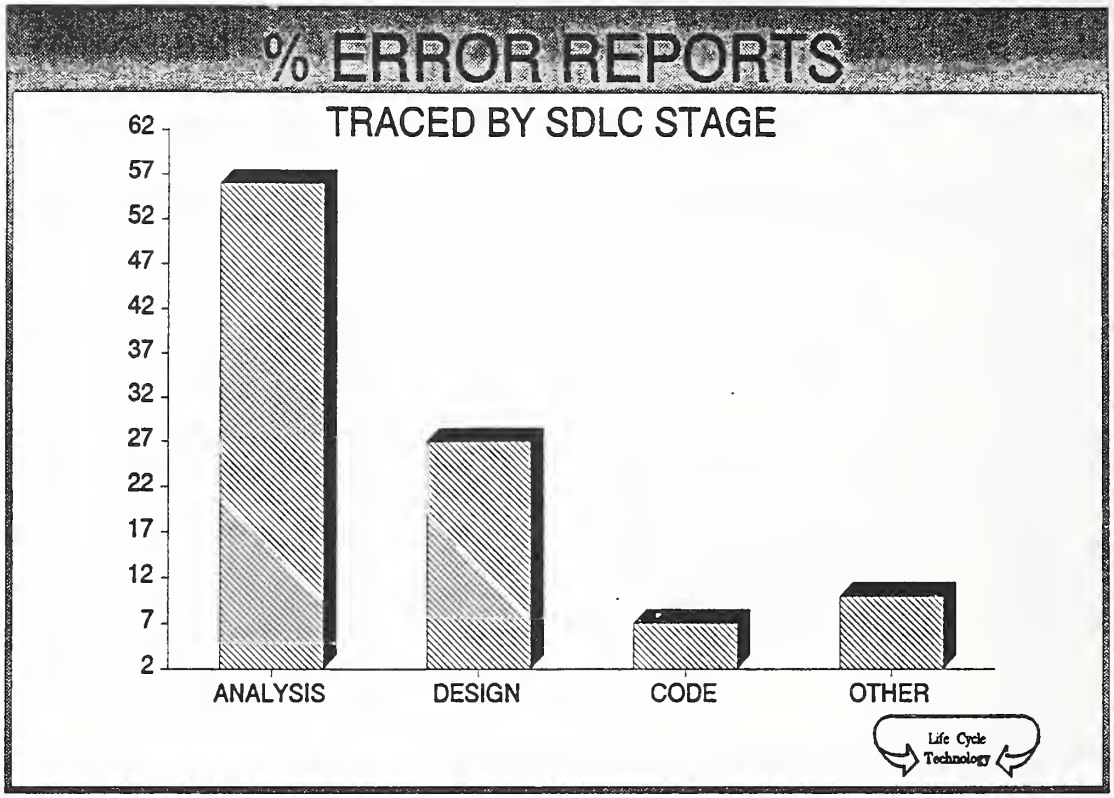
- Today's Systems Are Much More Complex
- Assume Study Off By an Order of Magnitude
  - Still Only 20% of Systems High Quality
- Recent Studies Show 20% to 30% of Large Systems Fail
- Systems That Require Extensive Rework Have Very High Operational Lifetime Costs
- CASE Viewed As Magic Bullet
- Most Analysts Don't Know Criteria for Quality Data Model

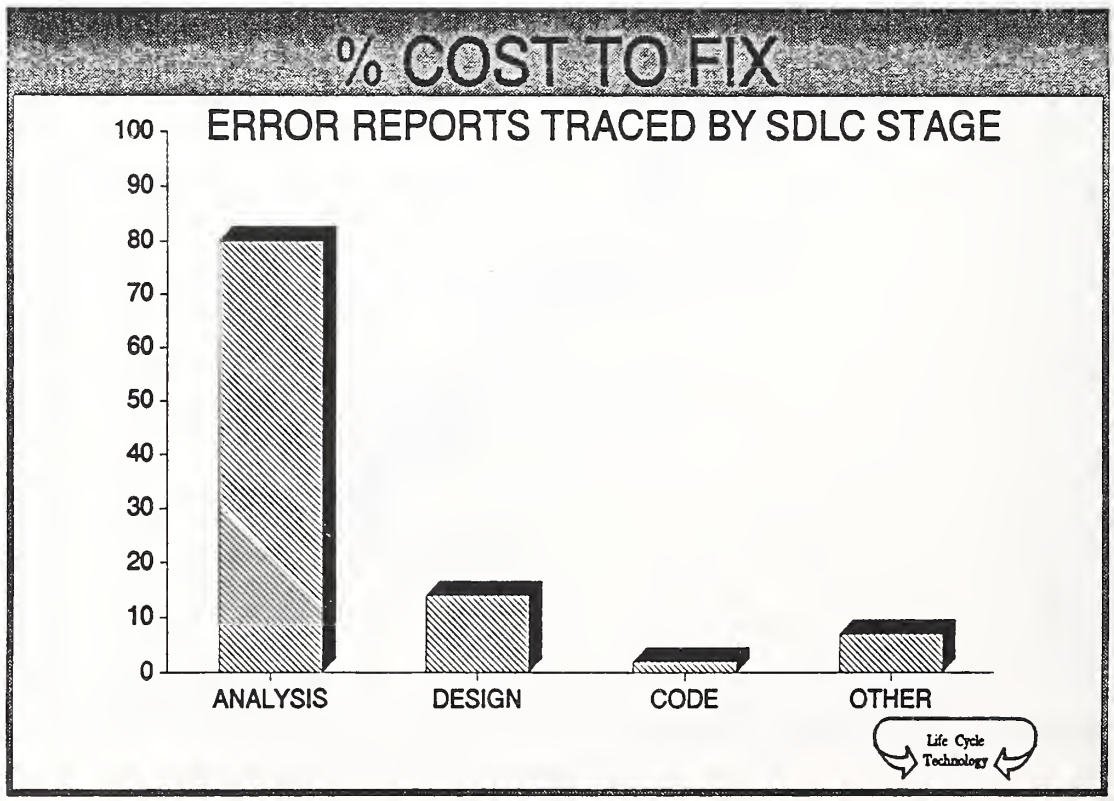
Life Cycle Technology

TYPICAL SDLC MODEL



Life Cycle Technology

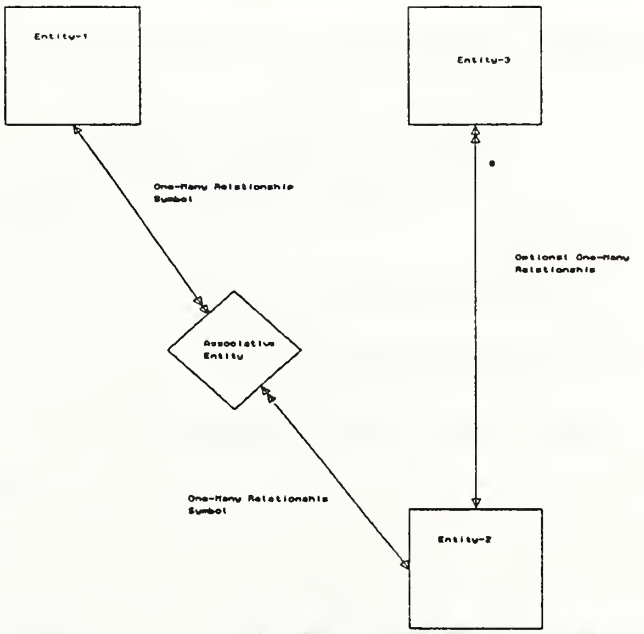

% ERROR REPORTS
TRACED BY SDLC STAGE
62
57
52
47
42
37
32
27
22
17
12
7
2
ANALYSIS
DESIGN
CODE
OTHER
Life Cycle Technology


% COST TO FIX
ERROR REPORTS TRACED BY SDLC STAGE
100
90
80
70
60
50
40
30
20
10
0
ANALYSIS
DESIGN
CODE
OTHER
Life Cycle Technology

EXAMPLES

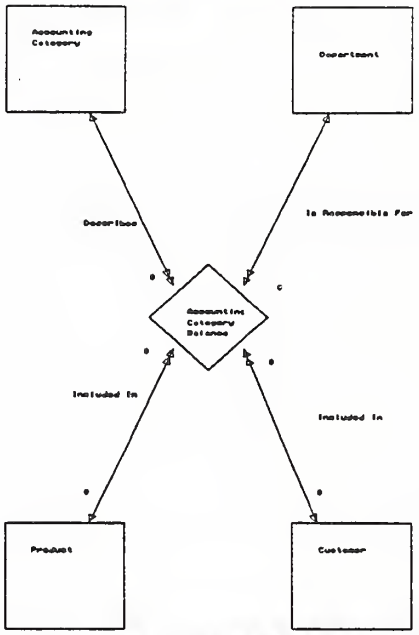
Life Cycle Technology

DATA MODELING SYMBOLS



Life Cycle Technology

ACCOUNTING EXAMPLE - TRADITIONAL DEVELOPER VERSION



Life Cycle Technology

ACCOUNTING EXAMPLE - TRADITIONAL DEVELOPER VIEW

ENTITIES

- Accounting Category Balance
  - General Ledger Data
  - Cost Accounting Data (Expense, Revenue, Statistics, etc.)
  - Balance Sheet Data
  - Customer Revenue and Expense Data
  - Product Revenue and Expense Data
- Accounting Category (Chart of Accounts, etc.)
- Customer
- Product

Life Cycle Technology

ACCOUNTING EXAMPLE - TRADITIONAL DEVELOPER VIEW

SYMPTOMS OBSERVED

- Extremely Complex Code
- Information Not Trusted
- Double Counting
- Processing Flags In Entities
- Slow Processing
- Change For One Area Caused Bugs In Other Areas

ACCOUNTING EXAMPLE - TRADITIONAL DEVELOPER VIEW

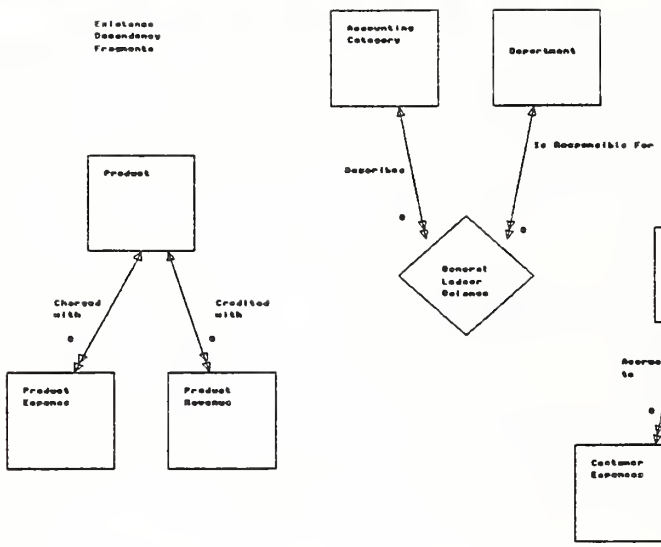
DATA MODELING ERRORS

- Combines Several Entity Types Into Accounting Category Balance
  - Based on Attributes With Similar Names

HOWEVER, DID NOT TAKE INTO ACCOUNT

- Data At Different Levels Of Aggregation
  - Customer and Product Revenue and Expense Data Already Summarized In GL Data
  - Cost Accounting Data Derived From GL Data
- Only Small Number Of Entries Pertained to Customer or Product
- Null Surrogates As Part Of Identifier
- Processing Flags As Attributes

ACCOUNTING EXAMPLE - DATA ADMINISTRATOR VIEW



Life Cycle Technology

ACCOUNTING EXAMPLE - DATA ADMINISTRATOR VIEW

ENTITIES

- General Ledger Balance
- Other Accounting Balances Not Shown
- Accounting Category (Chart of Accounts, etc.)
- Customer
- Customer Revenue
- Customer Expense
- Product
- Product Revenue
- Product Expense

Life Cycle Technology

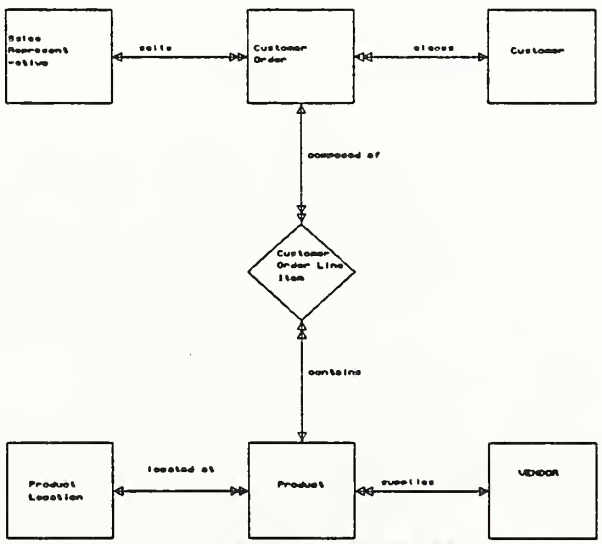
ACCOUNTING EXAMPLE - DATA ADMINISTRATOR VIEW

RESULTS

- All Data Within Any Entity At Same Level Of Aggregation
- All Attributes Completely Dependent On Identifier
- Clear Semantics
- Minimum Complexity
- Information Resource Integrity and Consistency

Life Cycle Technology

SALES EXAMPLE - TRADITIONAL DEVELOPER VIEW



Life Cycle Technology

SALES EXAMPLE - TRADITIONAL DEVELOPER VIEW

ENTITIES OF INTEREST

- Customer
- Vendor

MANAGEMENT GOALS

- Grow By Expanding Business With Existing Customers
- Buy From Companies Who Buy Our Products
- Build LongTerm Business Relationships

Life Cycle Technology

SALES EXAMPLE - TRADITIONAL DEVELOPER VIEW

SYMPTOMS OBSERVED

- Customers Leaving
- IS Reports Indicate No Problems
- Companies That Are Both Vendors and Customers:
  - Listed By Different Names in the Customer and Vendor Tables
  - Different Addresses in the Customer and Vendor Tables
- Unreliable Information
- Management Not Able to Implement Goals
- Salesmen and Managers Discharged

Life Cycle Technology

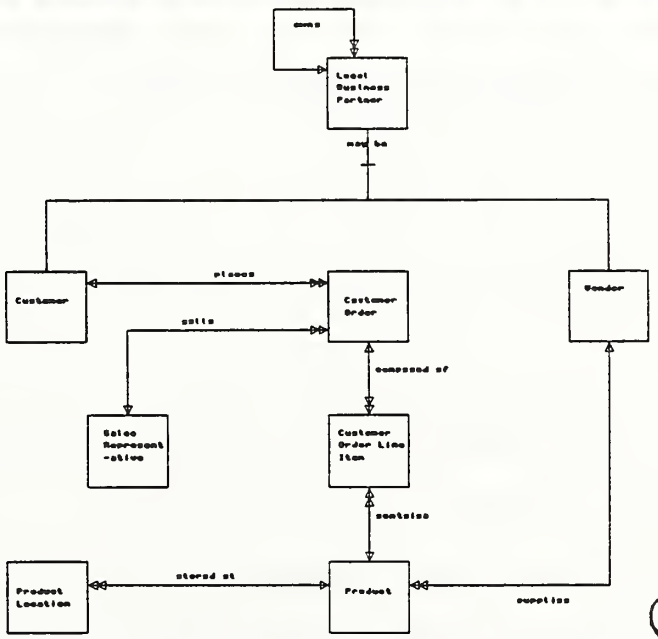
SALES EXAMPLE - TRADITIONAL DEVELOPER VIEW

DATA MODELING ERRORS

- Lack of Understanding that an Entity Can Have Multiple Business Roles
- Lack of Understanding that These Roles Have Information Overlap
- Duplication of Information

Life Cycle Technology

SALES EXAMPLE - DATA ADMINISTRATOR VIEW



Life Cycle Technology

SALES EXAMPLE - DATA ADMINISTRATOR VIEW

ENTITIES OF INTEREST

- Supertype Entity Business Partner
- Subtype Entities for Business Partner Roles
  - Customer
  - Vendor

Life Cycle Technology

SALES EXAMPLE - DATA ADMINISTRATOR VIEW

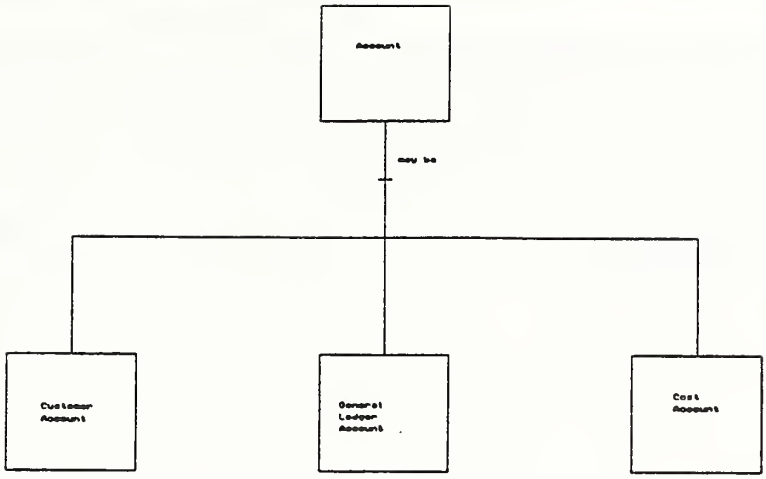
RESULTS

- Explicitly Accommodate Multiple Entity Roles
- Manage Role Information Overlap
- Eliminate Duplication of Information
- Information Resource Integrity and Consistency

- Management Better Able to Achieve Goals of:
  - Grow By Expanding Business With Existing Customers
  - Buy From Companies Who Buy Our Products
  - Build LongTerm Business Relationships

Life Cycle Technology

ACCOUNT EXAMPLE : TRADITIONAL DEVELOPER VIEW



Life Cycle Technology

ACCOUNT EXAMPLE : TRADITIONAL DEVELOPER VIEW

ENTITIES OF INTEREST

- Account(ing Categories)
- Customer Account
- General Ledger Account
- Cost Account
- Others Not Shown

Life Cycle Technology

ACCOUNT EXAMPLE TRADITIONAL DEVELOPER VIEW

SYMPTOMS OBSERVED

- Complex Integrity Rules
- Integrity Rules Hidden in Process Specifications
- Update Anomalies
- Complex Code

Life Cycle Technology

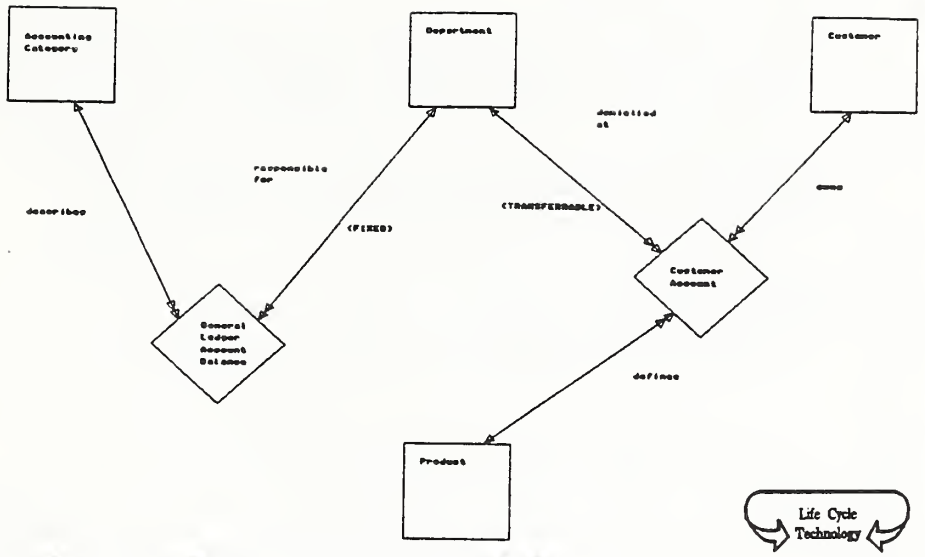
ACCOUNT EXAMPLE TRADITIONAL DEVELOPER VIEW

DATA MODELING ERRORS

- "Subtype" Real Existence Dependencies Not Shown
- Difference In "Subtype" Entity Life Cycles Not Apparent
- Different Rules for Creation Not Apparent
- "Subtypes" Are at Different Levels of Aggregation
- Homonym Mistaken for Supertype

Life Cycle Technology

ACCOUNT EXAMPLE - DATA ADMINISTRATOR VIEW



ACCOUNT EXAMPLE - DATA ADMINISTRATOR VIEW 2

ENTITIES OF INTEREST

- Accounting Categories
- Customer Account
- General Ledger Account Balance
- Cost Account Balance and Others Not Shown

Life Cycle Technology

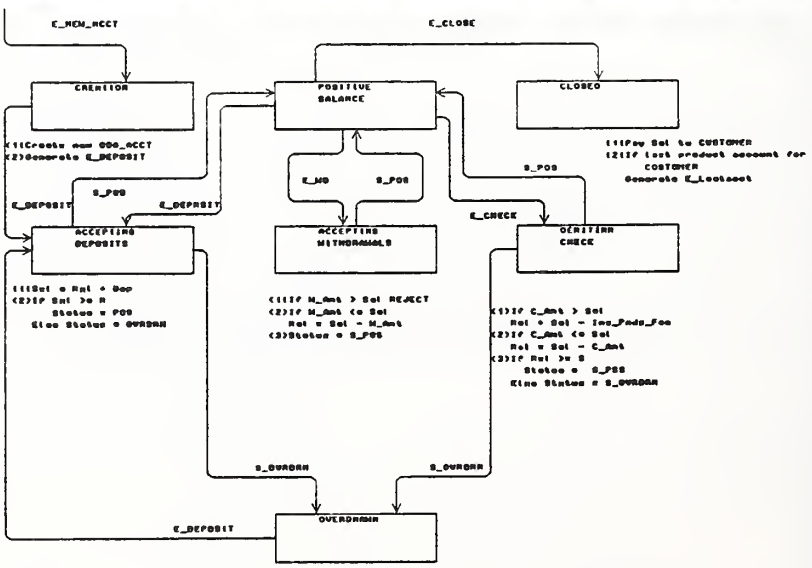
ACCOUNT EXAMPLE: DATA ADMINISTRATOR VIEW

RESULTS

- Existence Dependency
  - Accounting Balances Dependent On Accounting Category and Department
  - Customer Account Dependent On Customer, Product and Department
  - Only Existence Dependency in Common is Relationship With Department
    Customer Account Is Transferrable Relationship
    General Ledger Account Balance Is Fixed Relationship
- Entity Life Cycle
  - General Ledger Account Is Created and Archived
  - Customer Account Has More Complex Life Cycle
- All Entity Instances At Same Level of Aggregation
- Not Supertype/Subtype Structure At All

Life Cycle Technology

ALMOST NEVER SEEN - ENTITY LIFE CYCLE MODEL



Life Cycle Technology

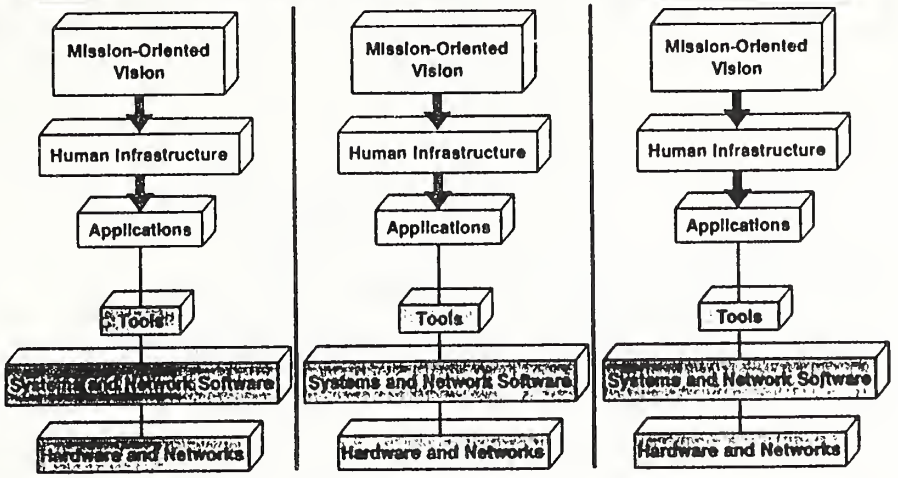
Thomas J. Buckholtz

Topic: Data Management and Information Proficiency: A Vision for the Future

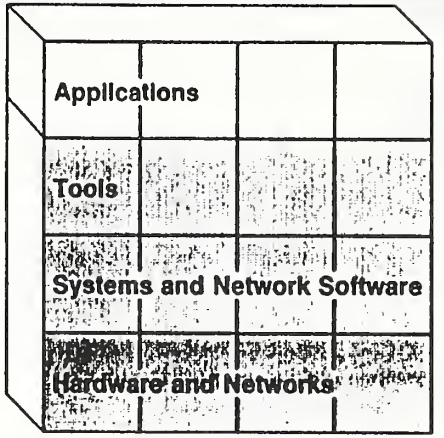
Description: Information proficiency is the effective use of information to accomplish an individual's job and an agency's mission. Essential to information proficiency is the successful management of data to maximize accessibility, accuracy or consistency, relevance, and strategic use. Vision, technology, people, organizations, and tools are critical elements in the equation of effective data management. Dr. Thomas Buckholtz will address these components of information proficiency, as well as his organization's program to accelerate information proficiency across all Federal agencies.

Biography: Thomas J. Buckholtz was appointed Commissioner of the General Services Administration's (GSA's) Information Resources Management Service in October 1989. He is responsible for government-wide programs dealing with the acquisition and management of telecommunications and computer equipment, systems and services. Dr. Buckholtz came to GSA from the Pacific Gas and Electric Company (PG&E) in San Francisco where he headed the company's Office Technology Project during the deployment of the firm's first ten thousand desktop computers. His work has included research, development and deployment of aerospace and defense systems, advanced approaches to energy production, computer applications, and new computers. He represents GSA on the Working Group on the Commercialization of Government Research, which supports the President's Council on Competitiveness. Dr. Buckholtz received an undergraduate degree from the California Institute of Technology and a Ph.D in physics from the University of California, Berkeley.

Traditional Paradigm



Modular Projects



Timeliness

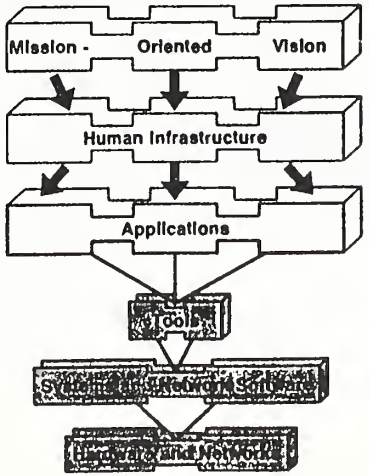
Lower Cost

Up-to-Date Technology

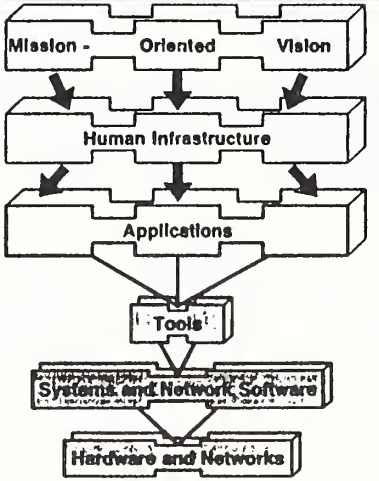
Flexibility

Less Risk

New Paradigm



New Paradigm

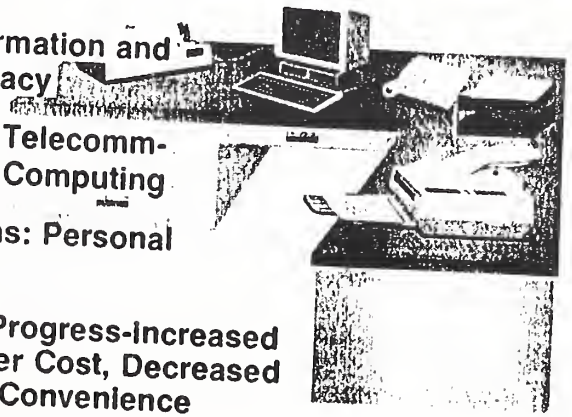


Information Architecture

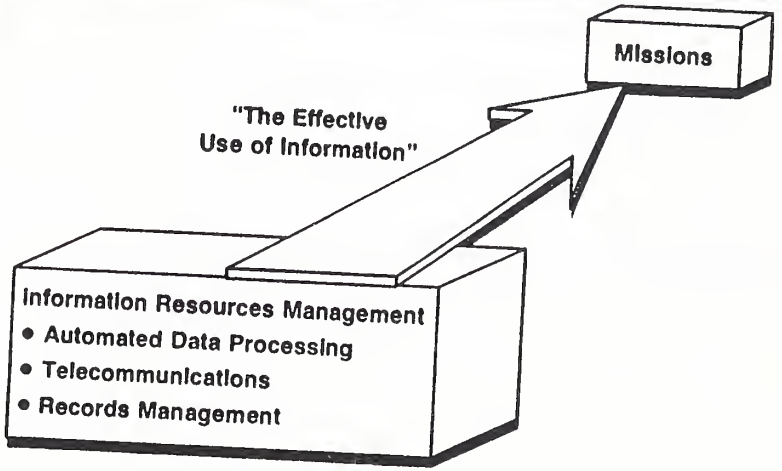
Technological Infrastructure

Evolving Workplace

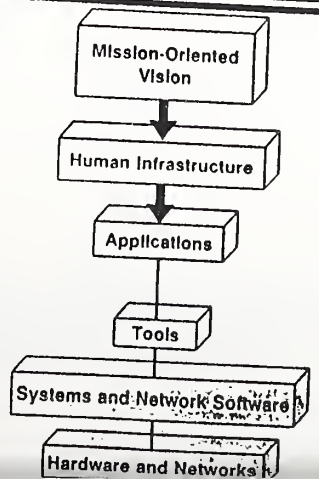
- New Applications
- Enhanced Information and Computer Literacy
- Covergence of Telecommunications and Computing
- Communications: Personal or Locational
- Technological Progress-increased Capability, Lower Cost, Decreased Size, and More Convenience



Information Proficiency

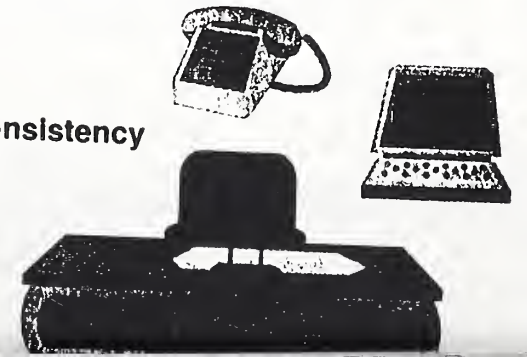


Architecture and Infrastructure



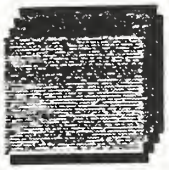
Workplace

- Space, Furniture, Equipment, and Supplies
- Colleagues
- Information:
  - Accuracy and Consistency
  - Completeness
  - Convenience



Services and Leadership

- Policy
- Telecommunications
- Technical Assistance
- Procurement
- IRM Services for GSA
- Direct Public Service



William R. Durell

Topic: Information Reengineering

Description: Mr. Durell of Data Administration, Inc. will present his views on information reengineering. In this presentation, he will provide a definition of information reengineering, and contrast this with software reengineering, forward and reverse reengineering and retrofitting. Unfortunately, most data reengineering efforts result in minor or cosmetic improvements to data quality. Cosmetic, reconstructive (severe) and moderate methods of reengineering will be presented. Mr. Durell will also discuss the various strategies employed to achieve reengineering, the relative risks and benefits associated with each strategy.

Biography: Bill Durell is the President of Data Administration, Inc., a consultant specializing in information resource management (IRM). He is the author of four textbooks and over twenty magazine articles on IRM. He conducts public seminars in major cities throughout the U.S. Mr. Durell has a Computer Science degree from Miami University and 20 years experience in data processing.

DATA REENGINEERING

WHAT IS DATA REENGINEERNG?

WHAT IS SOFTWARE REENGINEERING?

EXAMPLES:

- Reassessment or Reevaluation of Functional Requirements
- Object Oriented Programming
- Change to More Modern Programming Language
- Change From Batch to On-Line
- Instituting Programming Standards

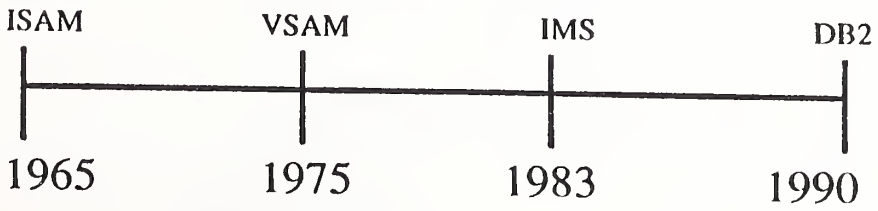
DATA REENGINEERING

- COSMETIC REENGINEERING
- RECONSTRUCTIVE REENGINEERING

CANNOT REENGINEER
INFORMATION UNLESS YOU
FIRST IDENTIFY THE PROBLEMS
YOU ARE TRYING TO SOLVE

TYPICAL SYSTEM REWRITES

TYPICAL DATA REENGINEERING

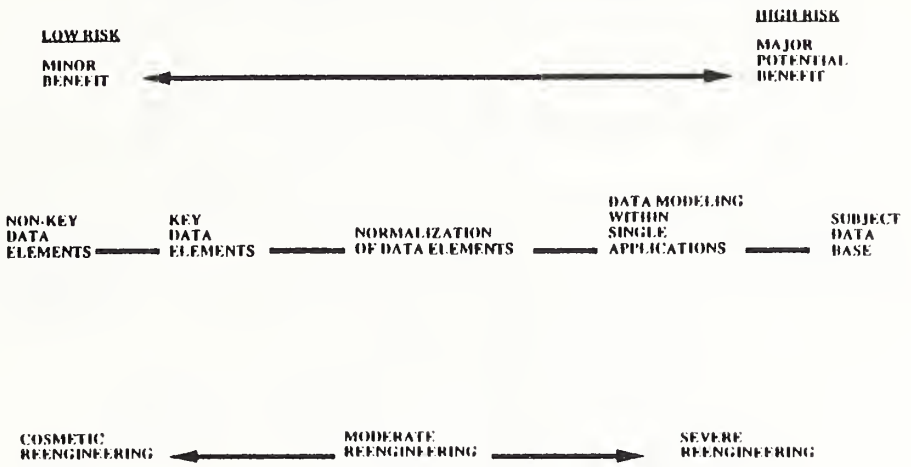


WHY WOULD YOU WANT TO REENGINEER DATA?

Existing Data-Related Problems For An Insurance Company:

1. **Cannot cross-sell products**
2. **Cannot determine customer profitability**
3. **Cannot determine customer population**
4. **No accurate demographics or psychographics**
5. **Cannot produce consolidated statements**
6. **Exaggerated mailing expenses**
7. **Loss of customer goodwill**
8. **Cannot determine customer-to-customer relationships**

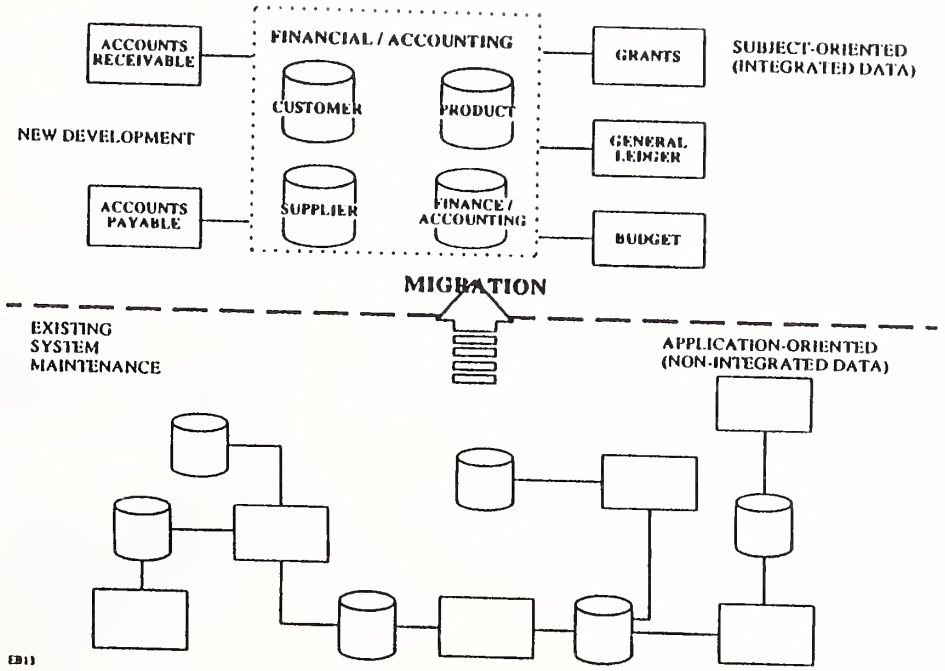
GRADIENTS OF DATA REENGINEERING



DATA MIGRATION

MIGRATION PLAN OR

CONVERSION STRATEGY



DATA MIGRATION

WHAT DOES MIGRATION MEAN?

WEBSTERS:

"The act of moving from one country, region, or place to settle in another".

QUESTIONS:

How Much Data Should Be Migrated?

How Much Data Can Be Mechanically Converted?

What About Historical Data?

DATA MIGRATION MISASSUMPTIONS

"UPLOADING"

"BACKLOADING"

"DOWNLOADING"

"SYNCHRONOUS UPDATES"

"PARALLEL FEEDS"

"MIRROR IMAGE FILES"

"PERIODIC COPYING"

"REFRESHING"

DATA MIGRATION VERSUS RETROFITTING

WEBSTERS:

RETROFITTING

"To furnish existing apparatus with new parts or equipment not available at the time of manufacture".

DATA RETROFITTING

"Modifying existing data files using data improvement or information engineering techniques".

MANAGING EXPECTATIONS

PRESCRIPTION FOR FAILURE #1:

1) Reengineer today's data to substantially improve its quality and eliminate as much redundancy as possible.

2) The reengineered data must be compatible with current files so that the two sets of data can be updated in a synchronized fashion until we can eventually wean the organization from the old data files.

MANAGING EXPECTATIONS

PRESCRIPTION FOR FAILURE #2:

1) Reengineer today's data to substantially improve its quality and eliminate as much redundancy as possible.

2) Mechanically or programmatically convert data in existing files to the new SDB.

MANAGING EXPECTATIONS

PRESCRIPTION FOR FAILURE #3:

"We like the idea of a SDB, so go ahead and architect and construct it. However, when we populate and update this SDB, we must hook or link this SDB with all similar data in existing files. After all, we need to keep the new data consistent with the old".

Graham W. Thompson

Topic: Managing Information Across Multiple CASE Tools

Description: The open tool architecture concept within AD/Cycle calls for the sharing of information via the repository between your organization's tools of today and tomorrow. This presentation addresses the practicality of this vision and the necessary steps for individual organizations and the industry as a whole to make this concept a reality.

Biography: Graham Thompson is National Sales Manager for Manager Software Products (MSP), Inc., North American Operations. Based in the company's Lexington Headquarters, Mr. Thompson is responsible for all MSP sales functions in the United States and Canada. Previously, he was Sales manager of MSP's United Kingdom Division.

Throughout his career, Mr. Thompson has played an active role in building application development environments for several large European companies seeking to integrate CASE tools through a central repository. While at MSP, Mr. Thompson has helped provide data management, DB2 and reverse engineering solutions for Europe's major software clients, including British Telecom, Europe's largest IBM shop.

Prior to joining MSP, Mr. Thompson worked at Shell Corporation as a business and systems analyst, before joining the sales and marketing division, where he helped implement the restructuring of Shell sales operations.

Graham Thompson holds a BS degree in Computer Science from the Portsmouth Polytechnic Institute in Portsmouth, England. He has given presentations worldwide on key issues in the program product/CASE area, including "The Importance of Strategic Information Planning" and "AD/Cycle: Repository Driven Application development.

(We were unable to include Mr. Graham's material.)

Bill Inmon

Topic: The Atomic Database: Building the Perfect Beast

Description: Five years ago, data architecture was an interesting idea. Three years ago, there were perhaps five companies in the process of implementing an architected approach to data. Today, there are many companies either in the process of completing their first thrusts into data architecture or in actually using their first implementation. Mr. Inmon will discuss some success stories, especially as those success stories relate to the business of an enterprise. In addition, some lessons learned along the way will be discussed.

Biography: Bill Inmon, with American Management Systems (AMS) in Lakewood, Colorado, is a writer and consultant specializing in data architecture, data base design, design review, and a variety of other topics. Of his 20 books in print, 11 have been book club selections. Mr. Inmon's books have been translated into Russian, Japanese, and Chinese. Mr. Inmon publishes his books and videotapes with QED, Wellesley, Massachusetts. In addition, Mr. Inmon is a columnist in DATABASE PROGRAMMING AND DESIGN and a frequent contributor to Enterprise Systems Journal. Among other contributions made to the profession of information processing have been the formalization of the notion of data architecture, the movement toward atomic data, the standard work unit, the maximum transaction arrival rate, the identification of the DSS parameters of success, and the notion of multiple levels of granularity.

In the '60s and '70s, coming from a world of master files, the notion of database was a single source of data for all processing. But in the '90s, with end-user computing, 4GL technology and massive amounts of data and detritus that have collected as a byproduct of on-line, transaction processing, the antiquated notion of database as a single source for all processing no longer suffices. There simply is too much data and too much diversity of processing to operate from a single database.

The evolution that is occurring (and in some places has occurred) in database is toward a dual database scenario. One type of database is for high performance, on-line transaction processing and the other type of database is for analytical, decision-making processing. On-line, high-performance databases have been around for years. It is the new type of database, the atomic database, that is of the greatest interest because it is just now that these new types of databases are being built.

The Atomic Environment

To best explain the atomic database environment that is evolving today, consider the following scenario. The boss walks in and asks that 10 years worth of customer activity be built into the database environment. The record of 10 years worth of customer activity will contain a large volume of data. In addition, the probability of access to any given unit of data is low. Is 10 years worth of customer activity deposited in the middle of an on-line, high-performance system? Of course not. To do so would be ruinous to the carefully-crafted on-line, high-performance environment. Instead, create an atomic environment, or an information warehouse. The term "atomic" stems from the fine granularity of atomic data.

Atomic/Operational Differences

The atomic environment is profoundly different from the classical operational environment. Some of the major differences are the following:

- Content - atomic data is made up mostly of historical data (archival data) and operational data is made up of current value data, for the most part.

- Manipulation - atomic processing consists of loading and accessing data with lots of calculation; operational processing consists of updating individual records.

- Record orientation - atomic processing usually operates one data set at a time; operational processing operates one or a few data records at a time.

- Organization - atomic data is organized into subject areas; operational data is organized application by application.

- Status - atomic data is just being organized and built, for the most part; operational data has already been built for the most part.

- Modeling - atomic data is oriented to a data model; operational data is oriented to a process model.

- Audience - atomic data serves a managerial audience; operational data serves a clerical audience.

- Usage - atomic data serves as a basis for heuristic, analytical processing; operational data serves as a basis for repetitive, detailed processing used to run the day-to-day activities of the corporation.

- Requirements - processing requirements are an unknown for much of the processing that occurs at the atomic level; processing requirements for the operational environment are stable and are known before the operational environment can be constructed.

- Mode - atomic processing consists of long sequential scans or loads of data; operational processing often consists of direct access and/or update of a few records.

- Integration - atomic data is integrated across the corporation; operational data is integrated only within the application that is being built.

The differences between the atomic and the operational database environments are so large that the two environments are almost alien to each other. It is a wonder anyone ever thought a single database could hold both types of data and service both environments.

Design Issues

In many shops the atomic environment is just now evolving. Some of the relevant design issues for the building of the atomic environment are granularity of data, partitioning, subject orientation, time variancy, extract processing and the development life cycle.

While granularity of data has always been an issue, in the operational environment the level of granularity was, for the most part, taken for granted. However, in the atomic environment granularity of data is the single most important design issue. Granularity of data has a profound effect on the volume of data and the usefulness of data that will be maintained. Indirectly then, granularity affects machine resources, response time and so forth. As a rule, if the designer handles granularity of data properly in the atomic environment, everything else falls into place. If the designer does not handle

granularity well, many other issues become difficult. Because of the volume of data found in the atomic environment and the unknown processing requirements that will be levied against the environment, it is mandatory that data be organized into partitions that can be independently accessed and indexed. For this reason, the partitioning strategy becomes the second most important design decision after granularity has been decided.

Designing atomic databases is aesthetically pleasing to the DBA who at long last can build subject databases. In an operational world dominated by systems that have already been built one application at a time, the DBA has faced a life of frustration. But with atomic databases (which are in front of the DBA, not behind him) and with no bias to one or the other application, the DBA organizes atomic data according to subject areas.

Each unit of atomic data has a moment in time associated with it (known as time variancy). It carries with it design considerations of complexity, volume, granularity and so forth.

Extract processing from the operational environments is a fairly complex procedure. The volume of data, the integration of data and the inability of the operational environment to support the extract all contribute to the complexity. While other aspects of design are typically overestimated, building the extract between operational and atomic data is usually underestimated.

The Development life cycle for atomic data is fundamentally different from the classical system development life cycle. Applying traditional methodologies to the atomic development process simply will not work. In addition, design techniques for atomic data must be tempered with the fact that atomic data is not updated. This has a profound effect on space management and physical database design.

There is a wide variety of other detailed design issues. The ones mentioned here provide only a glimpse of what is in store for the designer.

The Movement To Atomic Data

An interesting aspect of the movement toward atomic data is that it is widespread. The movement cuts across industries, size and technologies. Atomic Data is being built in financial institutions, insurance companies, manufacturers, service industries and so forth. There is no bias by industry for or against atomic data. In addition, atomic data is being built in large shops and small shops on a wide variety of hardware and software. In short, there is a sweep across the information processing industry toward atomic data.

Why is there a ground swell toward atomic data? Atomic data provides the basis to do something previously not possible -- to start to look at patterns of data or trends. Operational systems are good for the

management of individual records. But trying to detect major trends with operational systems is difficult.

And why is detecting trends or patterns so important? Because detecting patterns of processing, buying and so forth is at the heart of competitiveness in the marketplace. The sooner business trends can be spotted, the faster the business can react and capitalize on the trend. Without the capability of spotting trends, a business is at a serious disadvantage. This is why atomic data is the wave of the future in information processing.

Karen Lindsay

Topic: Data Management and Customer Satisfaction

Description: For a decade now, we have been evolving Data Management from an entity-relationship approach for data planning and data modeling, to the foundation for long-term, cost-effective, shared databases and decision support systems. We have introduced a number of tools to store modeling information, dictionary information, graphic data modeling information, etc., with some type of tool set. However, in evolving the corporate culture, providing customer satisfaction and delivering more than we promise is still more vision than reality.

The failure to achieve integrated, distributed, shared databases with consistent and accurate data should be viewed as only a short-term frustration. Success can be obtained through re-focusing on working with our customers versus around them. We must monitor our performance and continually be working to improve the process which will allow Data Administration and our customers to achieve long-term goals and avoid short-term frustrations.

Biography: Mrs. Lindsay is a Manager at Blue Cross and Blue Shield of Georgia, Information services Division. Her responsibility areas include data administration, data security, and strategic planning. Prior to Blue cross and Blue Shield of Georgia, she was Manager of Data Management, Contel Information Services, Coors Brewing, and a Data Architect of U.S. West Advanced Technologies. She has over ten years experience as a practitioner in data management. She is past Vice-President of DAMA, Atlanta, as well as the chapter founder.

BLUE CROSS AND BLUE SHIELD
OF GEORGIA, INC.

DATA MANAGEMENT PLANNING
CORPORATE ISSUES

- ACCOUNTABILITIES FOR PROCESSES AND DATA
- DATA CLASSIFICATION SCHEME
- INFORMATION ASSETS PROTECTION POLICY
- ARCHITECTURE GUIDELINES
- TOOLS AND CASE TECHNOLOGY IMPLEMENTATION
- METHODOLOGY

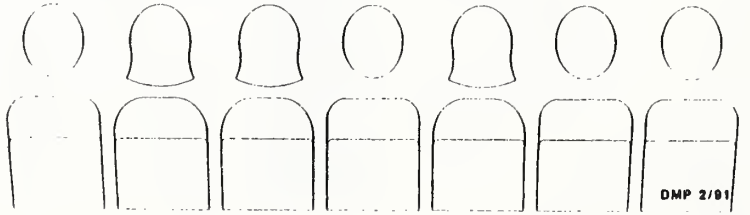
DMP 2/91

BLUE CROSS AND BLUE SHIELD
OF GEORGIA, INC.

DATA MANAGEMENT PLANNING

WHAT IS THE MESSAGE?

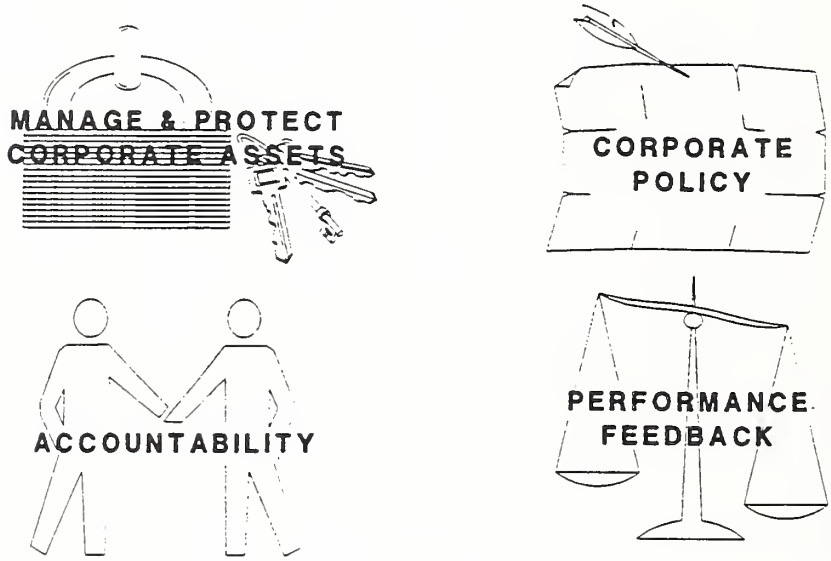
"DATA MANAGEMENT PLANNING
IS A BIG, IMPORTANT PROJECT.
WE MUST DEDICATE RESOURCES
TO IT TO MAKE IT HAPPEN."



DMP 2/91

BLUE CROSS AND BLUE SHIELD
OF GEORGIA, INC.

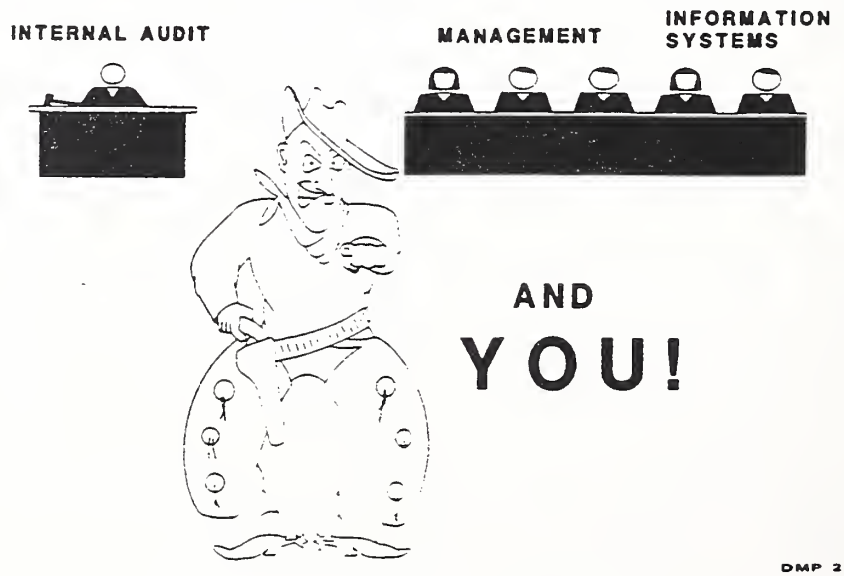
DATA MANAGEMENT PLANNING
PREDICTION



DMP 2/91

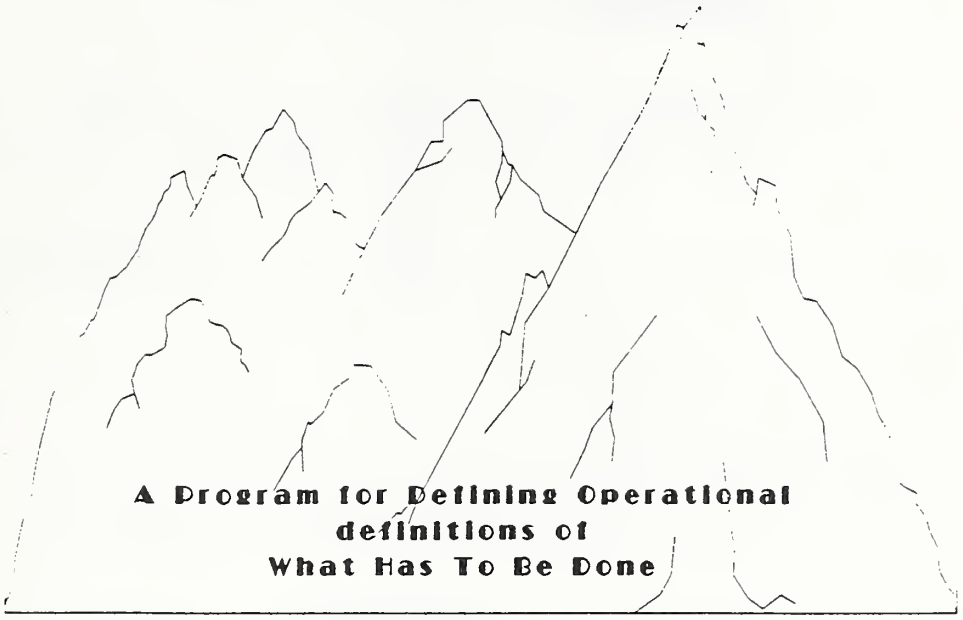
BLUE CROSS AND BLUE SHIELD
OF GEORGIA, INC.

DATA ADMINISTRATION PLANNING
WHO WILL BUILD THE PLAN?



DMP 2/91

CONQUERING THE DATA MOUNTAIN



A Program for Defining Operational definitions of What Has To Be Done

BLUE CROSS AND BLUE SHIELD OF GEORGIA, INC.

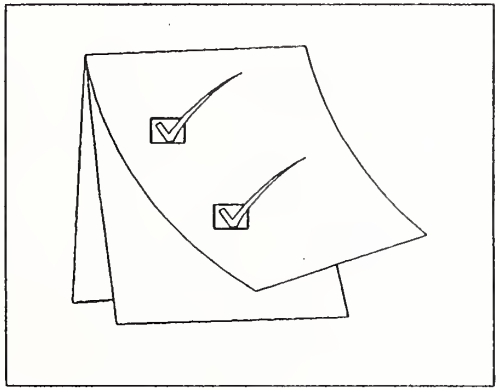
DATA MANAGEMENT PLANNING
PROJECT Versus PROGRAM

PROJECT
(One-time)

PROGRAM
(On-going)

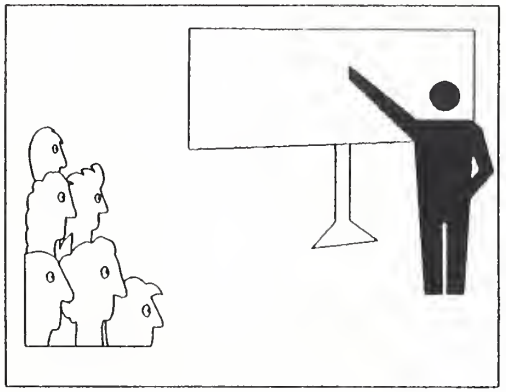
DMP 2/91

BLUE CROSS AND BLUE SHIELD OF GEORGIA
The Data Mountain
Customer Commitment Program



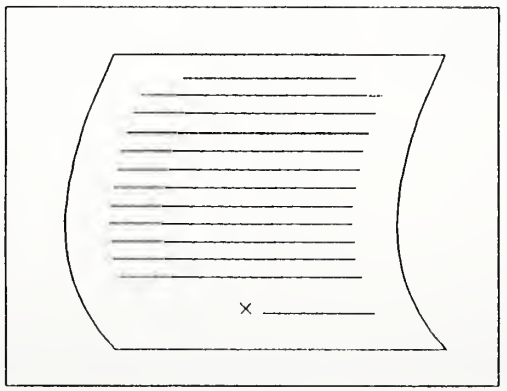
SURVEY

CUSTOMIZE



PROMOTE

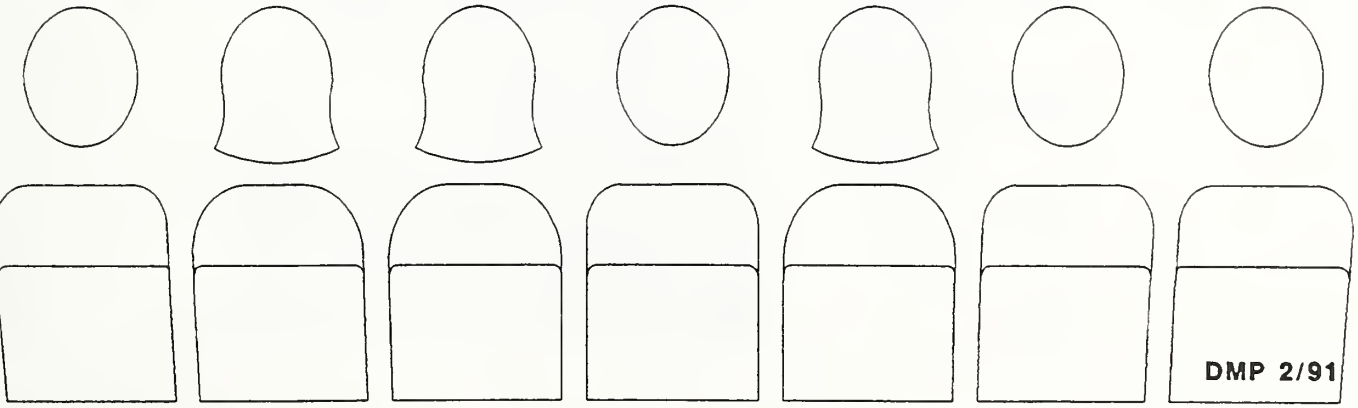
SERVICE



POLICY

BLUE CROSS AND BLUE SHIELD OF GEORGIA, INC.

DATA MANAGEMENT PLANNING DEFINITION & PURPOSE



INFORMATION SYSTEMS DEPARTMENT
DATA ADMINISTRATION
MISSION -- 1991

Optimize business information, through a corporate data perspective and data sharing

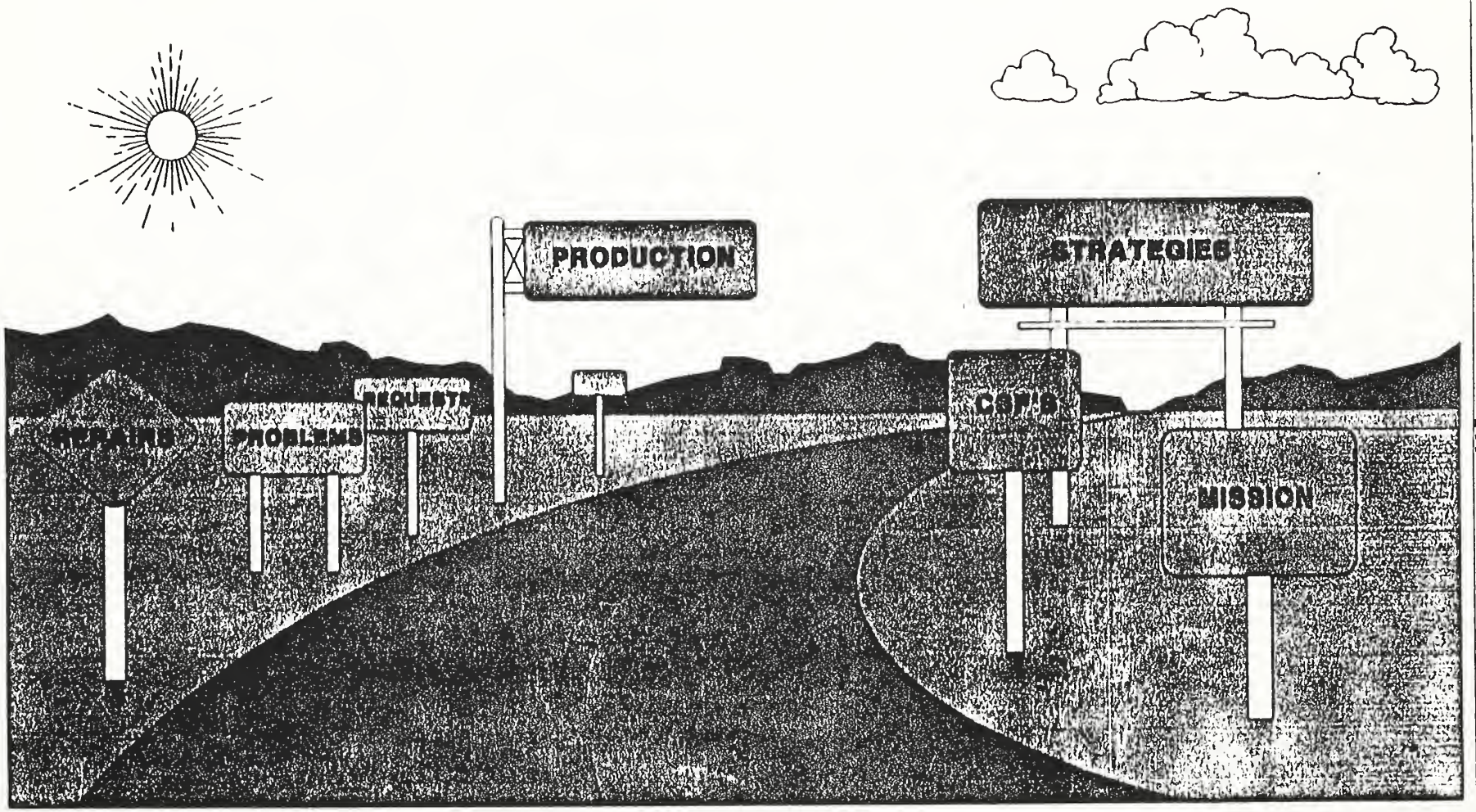
WHY ARE WE HERE?

BLUE CROSS AND BLUE SHIELD OF GEORGIA
The Data Mountain

Management Orientation
OBJECTIVIES

1. IDENTIFY THE BUSINESS REASONS FOR TAKING A MANAGED DATA APPROACH
2. DESCRIBE THE DATA ROLE IN PROMOTING THE CUSTOMER COMMITMENT PROGRAM
3. DEFINE THE CHANGE PROCESS
4. CRITIQUE THE CONSTRAINTS AND CHALLENGES OF GOING TO A DATA ENVIRONMENT CONFORMING TO MANAGED DATA
5. SCHEDULE IS CUSTOMER(S) FOR THE PROGRAM

BLUE CROSS AND BLUE SHIELD OF GEORGIA, INC.
INFORMATION SERVICES DEPARTMENT



Although we have to pay attention to the signs on the left, we want to be guided by the signs on the right.

12/89

Emmanuel Ackerman

Topic: Data Administration, the IBM Repository, and CASE Technology at Depository Trust Company

Description: Emmanuel Ackerman will present his view on the state of the IBM Repository in the short and long term. He will also present the Depository Trust Company's experience in working with the Repository and the stops they are taking to position for future releases. Prime among these has been an effort to implement an Application Development Methodology and Knowledgeware's IEW/ADW.

Biography: Emmanuel Ackerman has been working for the Depository Trust Company since 1978, managing data administration since 1984. Data administration is responsible for corporate data model, review of all data models for new or modified systems, and is currently implementing CASE technology. Mr. Ackerman is currently responsible for contingency planning technology assessment.

Data Administration, the IBM Repository and CASE Technology at DTC
Emmanuel Ackerman
May 15, 1991

DTC Overview

Depository Trust Company is a cooperative of the financial industry which was established to immobilize securities certificates and centralize some portions of post trade processing. DTC is currently custodian for over two-thirds of all nationally held securities, and settles over 95% of all trades made on the various U.S. exchanges. DTC employs over 3,000 people, 500 of whom are in data processing. Participants submit transactions and track securities positions and associated information over several on-line networks run off a single IBM 3090-500. IBM says that DTC is the largest DB2 implementation on a single CPU outside IBM in terms of quantity of data stored and volume of transactions processed. Total dasd on the production machine is around 400 gigabytes.

History of Data Administration at DTC

Patricia Graham, whom some of you may know, started Data Administration at DTC in the late 1970's. The group at that time handled all aspects of data administration from attempts at conceptual modeling to dasd space, pack back-ups and access methods. In 1984, I was transferred from the area supporting access methods to try to lead the floundering data modeling effort.

Our first activity was to look at the data dictionary market with an eye towards acquiring one. We selected DataManager, but were overruled by our management who had been talking to Codd and Date. They (our management and Codd and Date) felt that writing our own data dictionary on DB2 should be a straight forward and doable task. We did produce such a dictionary, which worked and was presented to at least two data administration conferences by members of my staff. The dictionary held all the important metadata, but lacked good performance since it was done in QMF, and lacked many of the inquiry capabilities which are necessary for fully supporting data administration and application development. I feel that the time we spent writing the dictionary would have been much better spent populating it, and leave the technical tasks to the technicians.

During that time, use of our modeling services was not mandatory for application developers. The DBA was slightly antagonistic towards us because she had tried to start the modeling effort and failed. She sent people to us who wanted a database, but hadn't defined what data they wanted in it. Consequently, we didn't have a lot to do. We created a corporate data model from corporate documentation, using interviews for follow up only. This received great acclaim within company management, and has led us to the development of several subject area databases. We also did a business area analysis of

reorganization functions. This did not directly lead to any changes in data or process in that area, but was used by the applications groups and end users as required reading for new employees, and was used to justify new people in the application development staff.

Shortly after that the DBA changed. The new DBA rapidly saw the difference between data structures we worked on and those we did not. The first production DB2 application was done by some consultants. It worked, but was unmaintainable and had to be rewritten within the following two years. The DBA established a rule that no one spoke to his people about new tables, columns, or views until they had approval from me. I was fully aware that applications groups work under deadlines and if I severely impacted their progress, they would simply go around me. Consequently, I do not try to review the business analysis or requirements. Rather, I have my staff insure that all data elements are atomic, cohesive, in conformance with naming standards... We require fully normalized tables, realizing that the DBA may denormalize for performance, but insuring that the denormalization is intentional, not haphazard. Also, we require justification of any tables which appear to be wholly or in part redundant. Again, only planned redundance for new data structures. There still is a lot of legacy data, and we are not perfect, we keep finding things that fall through the cracks. My boss, to whom the DBA also reports, and who was responsible for the successful implementation of DB2 at DTC even before I joined his staff, says that our effort is one of the key reasons DB2 has succeeded at DTC.

Repository Testing

Several years ago, the head of data processing at DTC was at a high level IBM briefing at which they pitched for the Repository. From that moment forward he instructed us to beat down IBM's doors to find out how DTC could wend its way to this new grand step forward. He selected four of us to lead the effort. We attended IBM briefings on the Repository whenever they were available prior to its initial availability. Over a year and a half before the initial availability, we heard the message to get on board with PS/2s, OS/2 and CASE. When we reported that back, we were told to go out and buy some. More on that later.

When the Repository came out we got it ASAP. Our testing included defining security as recommended by IBM, establishing some storage views and redoing class exercises on our system. Then we tried to load our corporate data model into the Repository. We never finished that task for two reasons. There was no particular usefulness in doing so, since no tool could retrieve it, and any future tool might have metadata requirements which could not be foreseen. Secondly, the Information Model is unbelievably complex and not fit for human consumption, i.e. I recommend only trying to load it through a tool.

We learned several lessons from our close encounter with the Repository. The Repository will be a critical vehicle in the glorious

view that AD/Cycle presents, but the way that companies such as ours prepare for that future awakening should be through implementing tools which will have an impact today, CASE tools and state of the art methodology. The information Model is not yet sufficiently robust to cover the application development cycle. The Repository itself is a performance dog. It is very interesting as an ER database, but it is not a modeling tool, and as I said before, it is not fit for human consumption. The CASE tools that are available today from IBM and its business partners will eventually communicate with the Repository and use it as the corporate coordinating repository. Until that day, the Repository is great for the tool builders of the world, but merely another overhead expense for the tool users. When the tools we can use successfully communicate with the Repository, we will be first in line to bring it back in.

Some of us were fortunate to go to very early Repository classes. These classes were not given frequently or locally so I ended up giving the classes from my notes. I made one major change when I presented the information. In trying to explain the various submodels, extensibility, and the differences between data, metadata, metametadata..., I found it far easier to start with an explanation of IRDS. I would preface the discussion with the statement that the Repository is not IRDS compliant, but it has some of the same conceptual foundation. Understanding IRDS is far easier than understanding the Information Model.

CASE Implementation

As I said in discussing the Repository, my group was dubbed with the responsibility of introducing CASE when we told management that introducing CASE would prepare us for the Repository. Upper management told us that there was money left over in that year's budget and we should go and buy some. Working in a very tight time frame it was easy to configure PS/2s and OS/2. The CASE tool selected was the market leader, but implementing an older methodology. We brought in some tool, classes and rapidly saw that we had made two major errors. There were far better tools on the market, and we needed to spend a great amount of time with the target application developers in order to get them to use the tools properly.

With our tails between our legs we went to upper management and requested to replace the previous CASE tool with Knowledgeware's IEW (and later ADW). We told them that while IEW was not perfect, there were not any tool sets better, and IEW best fit our corporate culture. Also we needed new staff to write a methodology and teach its use, or at least spread the message of analysis and design in a practical way.

We distributed the CASE tools through pilot methodology projects using an information engineering methodology. The pilot projects have been chosen as non-mission critical projects with a charter large enough to analyze an entire business area, and no immediate or hard deadline. We were pressed into doing a pilot before we were ready and of course

it failed. It ended in analysis paralysis facing an impatient user. The application area that we worked with on that first pilot liked what we were doing and immediately proposed another. We selected two pilots which were started last summer. One of them completed analysis in January and is now in design. That one should be fully implemented by the end of the year. The other one almost completed analysis around the same time. Various problems arose which had nothing to do with the project, but they eschewed any further work on the project until the fourth quarter. As of this writing we are discussing candidates for the next pilots, which will be started shortly.

One lesson of our first pilot was that without a direct commitment of end user time and availability, any pilot would fail. End user areas tend to be at first hesitant to get involved with the pilot projects because of the expected time involved and unknown territory. Once the users vent a little steam about the existing systems and hear what they do for a living, they get excited.

One of the two new pilots will be with the application area where the first one and one of the second two were done. The management of that group is so impressed with the results of BAA (without seeing what design might bring) that they want all the senior people in the area trained and doing that kind of work.

Our management likes the general approach we have taken. The man in charge of data processing (whom I mentioned before) has said, "This is like when we converted from assembler to COBOL, but this time we are moving the stop beyond COBOL. I had to force people to stop coding assembler and now they will need to learn new techniques." and "In ten to twenty years the applications programmers in this company will be out of work. All that will support, data administration and users."

The biggest issue we have faced is the cultural change. Programmers and analysts have a way of doing things that work. They say, "if it's not broken, don't fix it." We have been told not to claim that it is broken. (We have the same maintenance profile as the average shop does right now.) So we have to evangelize the gospel of quality and performance improvements through advanced methods. Trying to convince applications groups that there is another way to do analysis and design than whatever they might do today, is a tough fight. They figured out how to get the job done through a lot of sweat and toil, and now have a heavy investment in their way of doing things. As a consequence, we are starting to bring courses in house to address related subjects for those people not involved in pilots, such as structured programming, data modeling, structured analysis and design. Hopefully, exposure to these courses will help people understand what we are doing more easily. Industry experts suggest that it will take two to three years to implant a company-wide verbal methodology. We are intentionally trying to take that step.

We have been very development and upper management about how long the pilot projects will take from the very outset. At first they were slightly alarmed that our predictions about pacing were correct. But then they saw what was being done. Our management of their expectations has been important.

I have run into several frustrations using CASE tools in general. While most CASE tools support data modeling, the two which I have used do not do anything to help data administration do its job. They do not support naming standards checkers, they do not present the logical data information so that it is easily and quickly reviewable. They do not assist in creating meaningful names. They do not allow pruning and grafting of attributes. Administration of the information in the CASE tool can take special skills and is very time consuming. Bridges exist to data dictionaries, but those are bridges, not live connections. Most of these issues will be more easily resolved when the Repository becomes a reality.

The best automation models the way the best workers do their jobs. CASE automates application development but there are no products which model the best developers. All CASE products suggest that a person do one step or a small group of steps at a time, such as complete analysis before starting any design and coding. The best developers will start analysis, but on the first day of analysis they may do a smidgeon of design or write code fragments. No CASE tool supports that.

Bruce K. Rosen

Topic: Status and Application of Standards for Data Administration

Description: There is a vast number of data processing-related standards in the world today. Trying to identify and then sort through these standards, to know which standards to apply where, is often a difficult task. This task is further complicated by a common lack of knowledge about the current status of the different standards. This briefing will concentrate on identifying what standards exist that could be useful to data administrators, the current status of these identified standards, and will offer some suggestions on where these standards could be applied.

Biography: Mr. Rosen is the Manager, Data Administration Group, National Institute of Standards and Technology. Mr. Rosen received his degree in Information Systems Management and Computer Science from the University of Maryland. He then worked as a contractor performing real time control programming for NASA at the end of the Apollo project. Mr. Rosen later worked for the Department of Defense doing data dictionary and data administration programming systems analysis in the areas of Command, Control, and Intelligence. Mr. Rosen has served on several different American National Standards committees, including X3L8, the data representation committee, and X3H2, the committee that developed the NDL and SQL Database Language standards. He has been directly involved with he development of the Information Resource Dictionary (IRDS) since early in its development, both helping to define Federal Government IRDS requirements as a Data Administrator with the Defense Intelligence Agency, and now, as the Manager of the NIST Group responsible for the IRDS Federal Standard.

DATA ADMINISTRATION AND STANDARDS

Bruce K. Rosen
Manager, Data Administration Group
National Institute of Standards and Technology (NIST)

Environment

The data processing environment in which Data Administrators must operate is vastly different from that which existed only five years ago. No longer can organizations perform their data processing based on the old "input-process-output" model. Today's organization must function based on the concept that it is the data that is central to "data processing."

Problem

Information systems are constantly becoming more complex, costly, and important to the very existence of organizations. This importance to the organization cannot be overstated. According to a survey, only two out of ten companies whose data centers were destroyed were still in existence one year later [DURE85]. As these complex systems grow, the demands on Data Administration grow not only due to increases in data volume, but also because of other factors such as demands for management of non-traditional data and novel user interfaces. Some of the forms of non-traditional data that must be managed include graphics data, product production data, knowledge rules, text, etc. One example of a novel user interface is hypertext, which is capable of accessing many traditional and non-traditional data sources.

Given the problem, what can be done to solve it? New hardware and software will provide part of the solution, although these sometimes introduce new problems during the process of helping to solve existing ones. Since they increase system flexibility, interoperable components, based on standards, are also important elements of the solution.

Standard Interoperable Components

Using standards to develop interoperable components offers many advantages. Users will be able to protect their investments in data and programs. Vendors will have a large, established, and stable market which will also provide economies of scale.

However, it must be recognized that standards are not perfect solutions. By their very nature, standards are compromises. To help ensure acceptance they must offer some compatibility with existing products and data, and must be applicable to a variety of applications. However, without standards, anarchy rules. Thus, while standards may not be perfect, they are "enablers" of interoperability, not inhibitors.

Organizational Role of Data Administration

In order to organize the use of standards in Data Administration, it is necessary to first discuss a simple Data Administration reference model. This reference model, figure 1, serves two purposes in this paper. First, it provides a simple framework on which to base a discussion of the various information management roles performed to support the information needs of an enterprise. Second, it is used in the discussion of applicable standards to show where these standards would fit in the overall information management structure.

In this paper the information management reference model is divided into four parts: (1) Information Resource Management, (2) Data Administration, (3) Data Modeling Tools Administration, and (4) Database Administration. It is recognized that organizations may structure their information management components differently, and place emphases on other areas. Although the Software Engineering function is not addressed in this paper, it is essential that it be coordinated with the Data Administration function during the system lifecycle.

Information Resource Management (IRM)

IRM is the management function responsible for corporate level control of all information resource aspects of an organization. A variety of responsibilities may be managed under the IRM umbrella. Some IRM functions may include responsibility for computer software purchases, computer software development and maintenance, computer hardware purchases and maintenance, personnel resource assignments, etc. [ROSE89]

In making decisions concerning these different responsibilities, IRM must not only be concerned with high-level technical information resource issues, but also with a number of management issues involving the enterprise at large. For example, IRM must be concerned with implementing corporate goals, following corporate growth strategies, earning the expected return on investment, keeping expenditures within the limits of funding availability, training and retaining personnel, and keeping programs within projected limits for physical growth.

While high-level technical information resource issues must be involved in IRM decision-making, "real world" management decisions are often based on other concerns. Accordingly, higher level management decisions may sometimes be reached without much consideration for technical issues. Even issues such as personal influence and corporate "pecking order" can have a greater impact than technical issues on the decisions reached at this management level. IRM must operate in the area between high-level management's perspective of the enterprise as a whole and the technical operational perspective of managing the organization's information resources. IRM interprets high-level management's directives in terms of information resource management policy for Software Engineering and Data Administration.

IRM determines what applications and data are needed to support the organization's requirements. Software Engineering and Data Administration then determine how those applications are developed, and how data is stored, accessed, and maintained. Data Administration and Software Engineering also provide some technical advice on managing information resources to IRM.

Data Administration

Data Administration is concerned with the definition, control, and effective use of all information resources within an organization. One of Data Administration's primary areas of responsibility is establishing and maintaining the flow of information into, within, out of, and among the information systems used by an organization. [ROSE89]

Information collection, structure, storage, and access, even if not automated, are other areas of Data Administration responsibility. It is the responsibility of Data Administration to ensure that the data being collected and maintained within an organization is structured and stored in such a manner that it is accessible and understandable to all segments of the organization that have a legitimate interest in the information. To this end, with the cooperation of Software Engineering, Data Administration establishes and enforces organization-wide policies and standards involving data and system-related information.

Data Administration is responsible for ensuring that adequate records are kept to accurately reflect the "who, what, when, where, why, and how" of all data being collected, manipulated, or maintained within the organization. Finally, with the cooperation of Software Engineering, it is the responsibility of Data Administration to assign the available data system and storage resources to best serve the needs and priorities of the entire organization.

IRM is responsible for developing information resource policies used throughout the entire enterprise. Data Administration enforces those policies, and in turn, also directs the functions of Database Administration and Data Modeling Tools Administration.

Data Administration directs the function of Database Administration by overseeing a number of Database Administrators, each of whom is responsible for one or more particular databases. A database may be a database application implemented on a database management system (DBMS), or it may be a collection of "flat files" used together and collectively referred to as a database.

In the performance of Data Administration, automated data modeling tools are often used. Data Administration directs the function of Data Modeling Tools Administration by overseeing a number of Application Administrators and Dictionary Administrators. An Application Administrator is an individual who is responsible for

maintaining a particular data modeling application on a Computer Aided Software Engineering (CASE) tool. Similarly, a Dictionary Administrator is responsible for maintaining a particular dictionary application on a data dictionary system or repository tool. If it is more appropriate to the organization, however, Data Administration may oversee a number of Tool Administrators, each of whom is responsible for one or more software tools.

Data Modeling Tools Administration

While Data Administration is responsible for interpreting and enforcing information resource policies for the enterprise, Data Modeling Tools Administration is responsible for managing and carrying out those policies for individual data dictionaries, repositories and other applications of modeling tools. [ROSE89]

Data modeling tools are used to develop and describe databases, information systems, and information interchange within and among systems. Data Modeling Tools Administration directly supports the Data Administration process.

In organizations that have integrated data modeling applications, Data Modeling Tools Administration may be implemented on an application basis. In this case, Application Administrators and Dictionary Administrators should carry out data administration procedures and policies for each data modeling tool application.

In organizations that have multiple tools and a proliferation of data modeling applications, Data Modeling Tools Administration may be implemented on the basis of particular automated tools. In this case, individual Tool Administrators should carry out data administration policies for applications running on each data modeling tool. Each data modeling tool can be used to support multiple data modeling applications.

Database Administration

While Data Administration is responsible for interpreting and enforcing certain information resource policies for the enterprise as a whole, Database Administration is responsible for managing and carrying out those policies for individual subject databases. [ROSE89]

Database Administration is concerned with the efficient use of the resources that hold and make available the data resources of particular databases within an organization. This is the area responsible for controlling and maintaining the flow of data into, within, out of, and among each individual database.

Database Administrators physically control how, where, and in what manner, data is stored and maintained within each database. Each Database Administrator ensures that his or her database is operating

smoothly and optimally, with a physical database design that reflects the current state and usage of the database.

At the same time, Database Administrators provide information to Data Administration concerning organizational use of data within the subject database. Since they are in touch with database users, Database Administrators are responsible for recognizing when and how users' needs change, and if these changes require modification of the subject database. When users' needs evolve significantly, a Database Administrator reports these changing requirements to Data Administration. If users' needs change significantly, Data Administration may decide that one or more existing databases require redesign, or a new database is required.

Data Management Integration

The four organizational areas depicted in figure 1 create an interconnected hierarchical structure. The figure illustrates how these four areas are structured and interconnected with shared interface activities.

Figure 1 shows that IRM is primarily concerned with the long-range management view of the enterprise. To this end, IRM defines the necessary corporate information resources to meet the enterprise's strategic organizational goals. While IRM should be sensitive to technical issues brought up by Data Administration, IRM must adjust information resource management policy to follow the directives of corporate management.

Data Administration, situated between the long-range management view of IRM and the sometimes short-range technical view of information management implementation, provides a mid-range view of the enterprise. Following IRM directives, Data Administration defines the technical means necessary to meet corporate information objectives, brings technical considerations to the attention of IRM, and mediates between potentially conflicting technical and strategic goals. To maximize the benefit of data resources, Data Administration defines organization-wide policies for Database Administration and Data Modeling Tools Administration.

Data Modeling Tools Administration directly supports the Data Administration process, and maintains descriptive information on an enterprise's information, databases, information systems, and other information resources. Data modeling tools are used in the development and maintenance of databases and information systems. Data Modeling Tools Administration is concerned with ensuring that Data Administration users of data modeling tools are effectively and efficiently supported in their data modeling applications.

Database Administration maintains corporate information and data used in information systems. Database Administration is primarily concerned with making certain that database users are served

efficiently and effectively by the various components that make up a database, even if the Database Administration area does not itself directly provide the service. Database Administration follows the policies of Data Administration, bringing problems and necessary changes to the attention of the Data Administrator.

Organizational Interfaces

Figure 2, Organizational Interfaces, shows how the four aspects of the information management model interface with each other and with various other components of the organization.

At the top of figure 2, the high-level management of the organization sets the strategic goals for the enterprise and develops the corporate plans to achieve those strategic goals. From IRM, high-level management receives requirements for resource commitments and requests for policy guidance on corporate strategic goals.

IRM receives resource requirements from Data Administration, commits enterprise resources to corporate endeavors, and defines organizational policy for Data Administration. In turn, Data Administration provides organizational information resource requirements and policy to Database Administration and to Data Modeling Tools Administration. Data Administration is supported by Data Modeling Tools Administration and the tool set that it operates.

Since Data Administration enforces corporate-wide policy for information management, the various databases within the organization can be combined to form a corporate-wide integrated data resource. If implemented in the data modeling tool set, a data directory can be used to provide access to particular information within the integrated data resource. Similarly an information system directory can be implemented in the data modeling tool set, to provide access to particular systems and subsystems available within the organization.

Data Administration provides descriptive modeling information that defines both information systems and the integrated data resource. Instead of using fragmented and dispersed data resources, information systems utilize the corporate integrated data resource.

Data Administration interprets and enforces the policy that determines which components of the organization may access which aspects of the organization's data assets, and when new database assets will be developed. Thus Data Administration's interface with the other portions of the organization will be at the higher policy-setting levels of the organization. Information users come to Data Administration to request access to different areas of the organization's information assets or to request that new assets be developed.

Database Administration operates database management systems (DBMSs) and other software supporting databases that, in turn, support the

corporate integrated data resource. When corporate system users define their information requirements with the assistance of Data Administration, their local information requirements are handled by Database Administration, and their corporate information requirements are handled by Data Administration. Corporate system users access corporate information systems as part of the integrated data resource, which may be supported by one or more DBMSs.

In a sense, Database Administration serves as a valve that controls and directs the information flow between the actual physical databases of the organization, and the areas of the organization requesting access to that information. Also Database Administration is responsible for implementing any new database requirements defined by the organization. This situation results in Database Administration having a very direct interface with much of the working level of the organization in order to resolve the daily problems of accessing information.

Standardized Profiles

One of the most important activities in the data processing world today is the development of standardized profiles. These profiles provide the standards that organizations need to build their own data processing structures. One of the important aspects of these standardized profiles is that they cut across all data processing organizational lines in order to bring together the many different aspects of data processing. Two of the most important standardized profiles today are Government Open Systems Interconnection Profile (GOSIP), and Application Portability Profile (APP). For more information on these profiles the reader should refer to [FIP146] and [APP91].

SQL

From a Database Administration aspect, perhaps the most important standard today is SQL. SQL provides portability of database data definition and database application programs via standardized Data Definition Language (DDL), Data Manipulation Language (DML), and bindings to standard programming languages COBOL, FORTRAN, and Pascal. The current version of SQL was issued as a revised American National Standards Institute (ANSI) standard in 1989 (X3.135-1989 and X3.168-1989), and as a Federal Information Processing Standard in 1990 [FIP127]. Included in this latest version of SQL in enhanced integrity constraints and embedded SQL for Ada, FORTRAN, COBOL, and Pascal languages.

It is currently estimated that the next version of SQL, known as SQL2, will be finalized in 1992. SQL2 will be upward compatible to the current SQL. It will provide schema manipulation language for dynamic creation and modification of schema elements and "information schema" for run-time access to schema information. It will also include dynamic SQL for run-time creation of SQL statements and functional

enhancements such as: outerjoin, update and delete cascade, domains, date and time, case expression, data type casting, string operations, temporary tables, and general relaxation of previous language restrictions.

In 1995 it is expected that a third version of SQL, now referred to as SQL3, will be issued. This version of SQL is currently expected to include enhancements to support object-oriented database, knowledge-base systems, expert systems, and data administration.

Remote Database Access (RDA)

Remote Database Access (RDA), a standard presently in development, will provide protocols for remote access to databases distributed over a communications network. This standard will not provide distributed access or concurrency control. It is only a first step toward true distributed database.

Information Resource Dictionary System (IRDS)

The IRDS Standard is a set of software specifications for a standard data dictionary system. The IRDS Standard establishes the requirements for a software tool that can be used to describe, document, protect, control, and enhance the use of an organization's information resources. The description of information resources is a specific type of information referred to as metadata (i.e., data about data). The IRDS is intended to support the definition, management, and control of metadata. Specifically designed to support Data Administration as well as related Software Engineering procedures, the IRDS is the result of government and industry efforts to improve the functionality and utility of data dictionary systems.

The IRDS Standard was approved by the American National Standards Institute (ANSI) in 1988, and is published as ANSI Standard X3.138-1988. Following ANSI approval, the IRDS was adopted as Federal Information Processing Standard (FIPS) 156. The FIPS publication adopts the IRDS specifications provided in the ANSI publication. For a more complete discussion of the IRDS Standard, see [GOLD88]. For a detailed account of IRDS applications, see [LAW88].

Background of the IRDS Standard

The National Institute of Standards and Technology (NIST) has actively participated as a member of the technical subcommittee for Information Resource Dictionary Systems (X3H4). X3H4 is part of the X3 Accredited Standards Committee for Information Systems which operates under the procedures of ANSI.

In the early 1980's, NIST conducted surveys of Federal Government Data Administrators and other data management personnel to record their views of the limitations of the data dictionary systems then available. NIST also asked these same government workers about the

types of functionality they would ideally like to have in a data dictionary system.

With the results of those surveys in hand, NIST personnel devised a list of features that should be included in a data dictionary system to provide full support for Data Administration and related Software Engineering activities throughout the system lifecycle. This list of desired data dictionary system features [KONI81] was the starting point for the IRDS Standard.

IRDS as a Modeling Tool

The IRDS is designed to support information modeling through the information system lifecycle. Schema extensibility is an important feature of the IRDS that permits the expression of a wide variety of user-defined information models. This means that the dictionary administrator of an IRD application (running on an IRDS-compliant system) can define the schema for "types" of information to be stored in an IRD. It is up to the dictionary administrator of an IRD, and the IRDS users who will utilize that IRD, to define the appropriate framework, or IRD schema, into which to load the information about their organization and its resources. With this schema extensibility feature, IRDS compliant tools can satisfy the needs of many types of organizations.

Command Language, Panel, and Services Interfaces

The IRDS Standard currently specifies two types of user interfaces. The Command Language Interface permits IRD applications on the IRDS to be accessed through a set of procedural commands. The IRDS Command Language can be used either interactively or in batch mode.

The Panel Interface permits IRD applications on the IRDS to be accessed through a series of "panels" or screens. The Panel Interface is designed to provide the same full functionality as the Command Language. The Panel interface may be implemented through the use of screens, menus, and windowing.

A third type of interface, currently under development in X3H4, is the Services Interface. The Services Interface defines the needed functionality for communications between the IRDS and other software, such as a DBMS. This interface will provide the support required to use the IRDS in an "active" mode with a DBMS, or other types of Data Administration support tools, such as CASE tools. In an "active" mode, the IRDS will be able to interact directly with other software.

Export/Import Facility

There currently exists in the IRDS standard an IRDS Export/Import facility called the IRD-IRD Interface. At this time, the technical subcommittee responsible for the IRDS standard, X3H4, is working on finalizing the file format for this facility. This work should be

completed in 1991. The IRDS Export/Import File Format, specified with the Open Systems Interconnection (OSI) communications protocol Abstract Syntax Notation One (ASN.1), will be applicable to interchange among a variety of repositories, such as those utilized by CASE tools.

Central Repository

When the IRDS Services Interface is complete, and the IRDS can be used in an "active" mode, the IRDS can be used as a central repository in which metadata is stored for use by a variety of tools, such as CASE tools. These CASE tools can be used for support of graphical displays of information, while the IRDS can be used to store and coordinate the descriptive semantic information represented in CASE graphical displays.

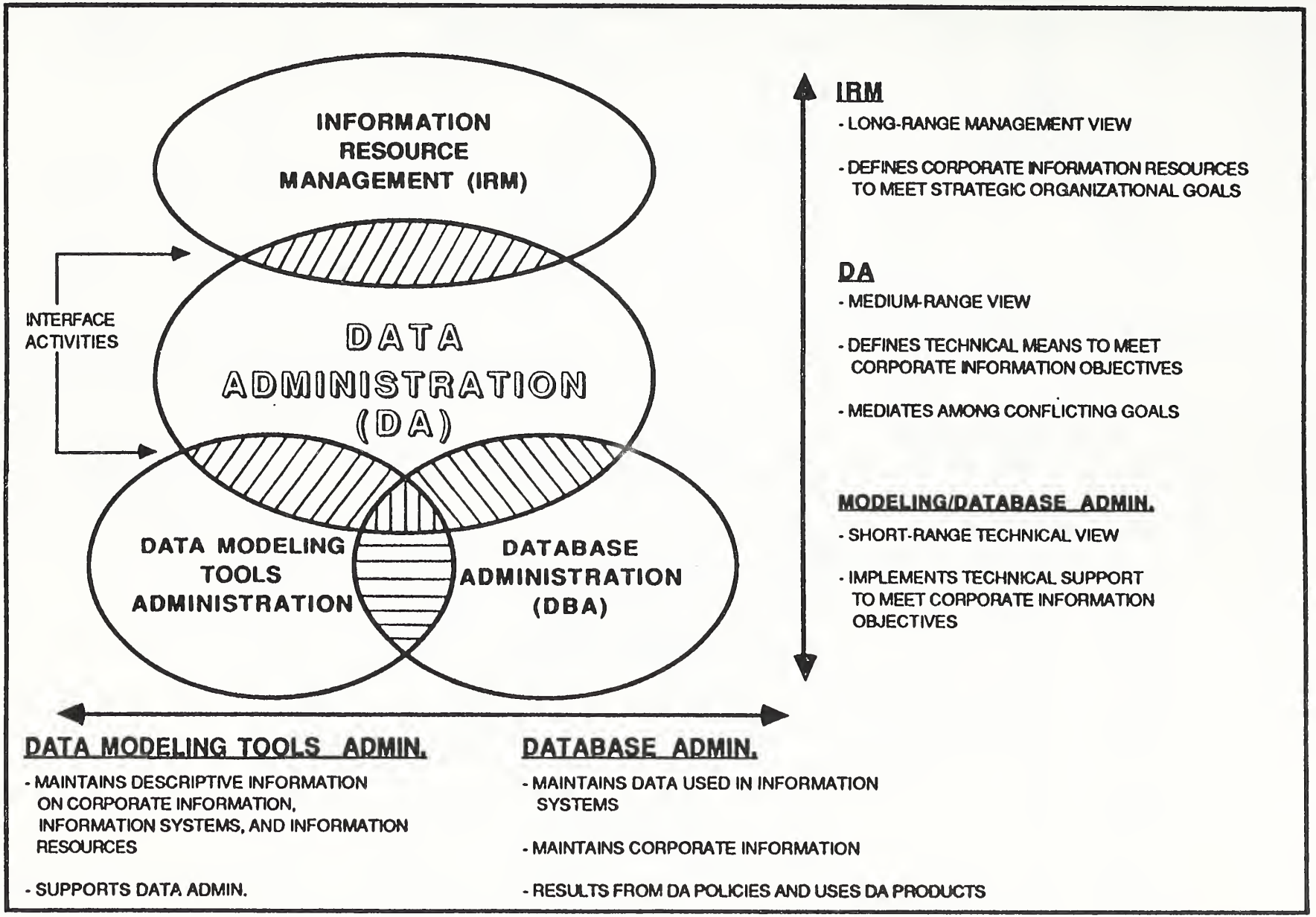

INFORMATION RESOURCE MANAGEMENT (IRM)
INTERFACE ACTIVITIES
DATA ADMINISTRATION (DA)
DATA MODELING TOOLS ADMINISTRATION
DATABASE ADMINISTRATION (DBA)
IRM
- LONG-RANGE MANAGEMENT VIEW
- DEFINES CORPORATE INFORMATION RESOURCES TO MEET STRATEGIC ORGANIZATIONAL GOALS
DA
- MEDIUM-RANGE VIEW
- DEFINES TECHNICAL MEANS TO MEET CORPORATE INFORMATION OBJECTIVES
- MEDIATES AMONG CONFLICTING GOALS
MODELING/DATABASE ADMIN.
- SHORT-RANGE TECHNICAL VIEW
- IMPLEMENTS TECHNICAL SUPPORT TO MEET CORPORATE INFORMATION OBJECTIVES
DATA MODELING TOOLS ADMIN.
- MAINTAINS DESCRIPTIVE INFORMATION ON CORPORATE INFORMATION, INFORMATION SYSTEMS, AND INFORMATION RESOURCES
- SUPPORTS DATA ADMIN.
DATABASE ADMIN.
- MAINTAINS DATA USED IN INFORMATION SYSTEMS
- MAINTAINS CORPORATE INFORMATION
- RESULTS FROM DA POLICIES AND USES DA PRODUCTS

FIGURE 1. DATA ADMINISTRATION REFERENCE MODEL

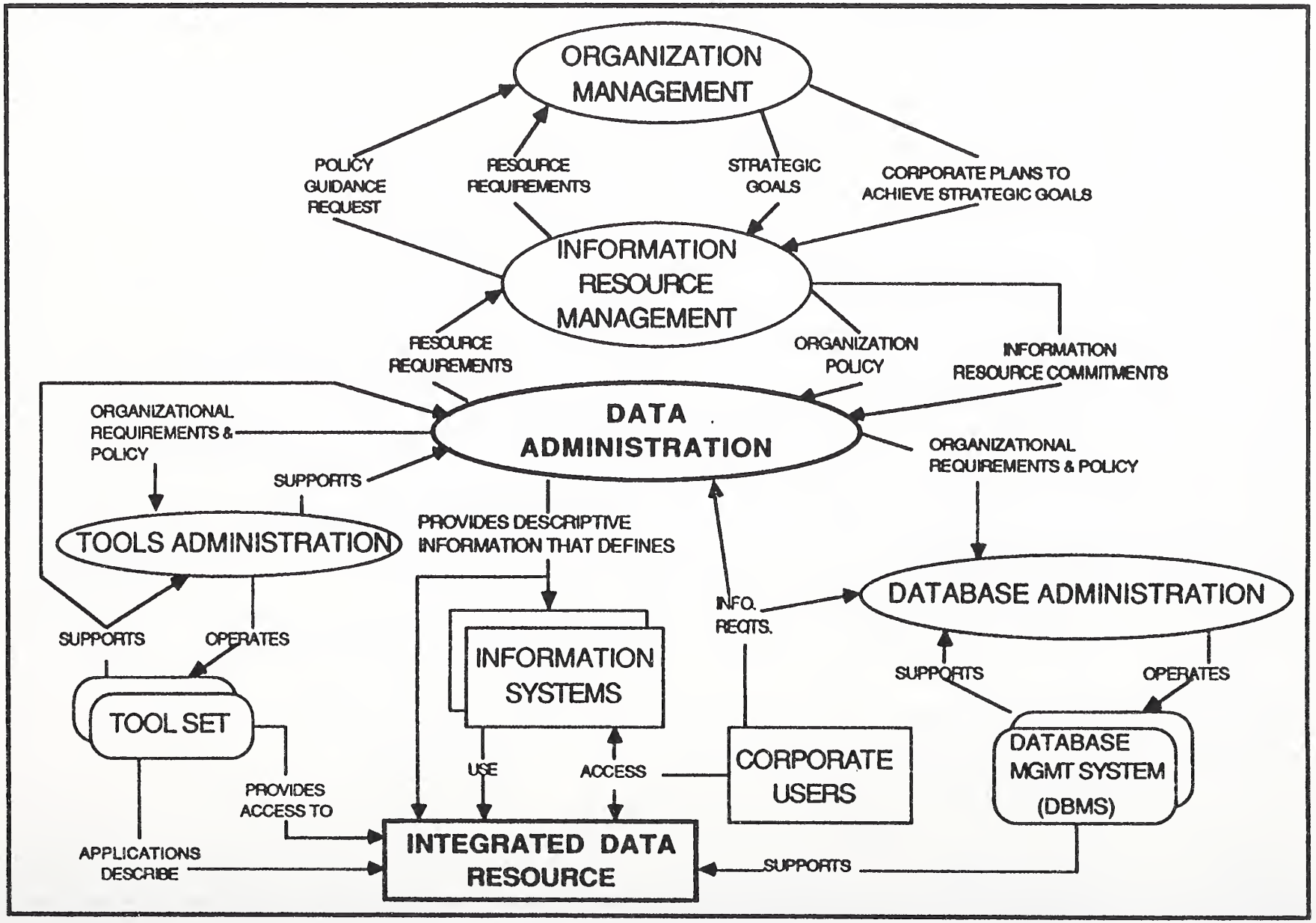


FIGURE 2. ORGANIZATIONAL INTERFACES

REFERENCES

[APP91] NIST Special Report, Application Portability Profile, The U.S. Government's Open System Environment Profile (DRAFT), National Institute of Standards and Technology, Gaithersburg, MD, January 1991.

[DURE85] Durell, William R., Data Administration, A Practical Guide to Successful Data Management, McGraw-Hill Book Company, New York, NY, 1985.

[FIP127] Database Language SQL, Federal Information Processing Standard Publication 127-1 (FIPSPUB127-1), National Institute of Standards and Technology, Gaithersburg, MD, February 1990.

[FIP146] Government Open Systems Interconnection Profile (GOSIP), 1988, Federal Information Processing Standard Publication 146 (FIPSPUB146), National Institute of Standards and Technology, Gaithersburg, MD, August 24, 1988.

[GOLD88] Goldfine, Alan, and Patricia Konig, A Technical Overview of the Information Resource Dictionary System (Second Edition), NBSIR 88-3700, National Institute of Standards and Technology, Gaithersburg, MD, January 1988.

[KONI81] Konig, Patricia A., and Judith J. Newton, Federal Requirements for a Federal Information Processing Standard Data Dictionary System, NBSIR 81-2354, National Institute of Standards and Technology, Gaithersburg, MD, September 1981.

[LAW88] Law, Margaret Henderson, Guide to Information Resource Dictionary System Applications: General Concepts and Strategic Systems Planning, NBS Special Publication 500-152, National Institute of Standards and Technology, Gaithersburg, MD, April 1988.

[ROSE89] Rosen, Bruce, and Margaret Law, Guide to Data Administration, NIST Special Publication 500-173, National Institute of Standards and Technology, Gaithersburg, MD, October 1989.

Reed Phillips

Topic: Future Vision: An Essential Prerequisite for a Data Management Program

Description: Developing and communicating a future vision of data management is a critical success factor for: strategic planning, securing top management support and commitment of resources, technology assessment and implementing the current year action program. Further, this vision or concept of the future must be constantly revised to respond to changing technology, management priorities and resource availability. Reed Phillips will address this and the importance of the current year program in realizing the organization's vision.

Biography: Reed Phillips has been Director of Information Resources Management (IRM) for the Department of Commerce since 1985. Prior to that, he was Vice President Systems Engineering for American Broadcasting Company (ABC). From 1980-1983, he was Director, IRM for the Department of the Interior. He was among the first inductees into the Government Computer News, Federal Government Information Technology Hall of Fame; was one of the top 100 CIO's in the United States in 1989 as recognized by CIO Magazine; and in 1990 was the recipient of the President's Award given by the Federation of Government Information Processing Councils.

INTRODUCTION

Vision is a difficult topic to introduce within the context of a discussion on technology or its management. It brings with it a connotation of escapism or unreality that seems inappropriate to the subject matter. However, many of you would agree that vision and management in the form of leadership have always been required to successfully resolve our more pervasive and difficult national problems. Many would also agree that the effective management of data has emerged as an issue of major importance to the information systems industry and the nation. If the government's massive information resources are to be used to address the problems of the environment, crime, education, public health and the economy, data management vision is needed. Researchers and others who can benefit in the public and private sectors must have access to the scientific and technical information required for America to compete in a global economy. Our vision and leadership must be equal to the task.

By way of response to this challenge, Secretary of Commerce Mosbacher has established as one of the Department's major program goals the difficult task to "provide data and information products and services to support governmental policy and meet public needs." One might consider this to be Commerce's main goal because Commerce, more than any other civilian federal agency, has programs and missions whose products are primarily information and information services. Success in achieving this goal will depend in large measure on the effectiveness of our data management initiatives and the future vision that will provide the driving force towards their accomplishment. Future vision is an essential prerequisite for a successful data management program.

The purposes of this presentation are to place data management vision in perspective, and to discuss differing visions of the future within the context of real world organizational and problem oriented environments. This will be accomplished by first defining what vision is and describing its components in terms of an entity relationship model. Use of the model to frame the discussion of data management vision is especially appropriate because of data's integral role associated with all aspects of the information systems life cycle and architecture. Then, examples of future vision in the public sector and private sector are presented as being representative of global vision statements that are indicative of the problems and issues in their respective communities. Finally, data management as it relates to major Department of Commerce missions is discussed together with vision statements that reflect individual bureau and department-wide perspectives.

MODEL OF DATA MANAGEMENT FUTURE VISION

MODEL OVERVIEW

The model of future vision (graphically displayed on the following page) identifies the component entities which contribute to it's formulation and the results that must occur to achieve success. **Business objectives, technological change, existing architectures**, and **realistic budget estimates** are the essential foundation ingredients in the creation of our vision. The **Information Systems (IS) Plan**, which is the formal means of communicating the vision to all concerned and interested parties, is a unique entity. It provides input to the development of our vision and is the primary output product together with it's associated budget request. The **communication** and **implementation** components of the model are action oriented entities that are essential in achieving the vision.

THE VISION

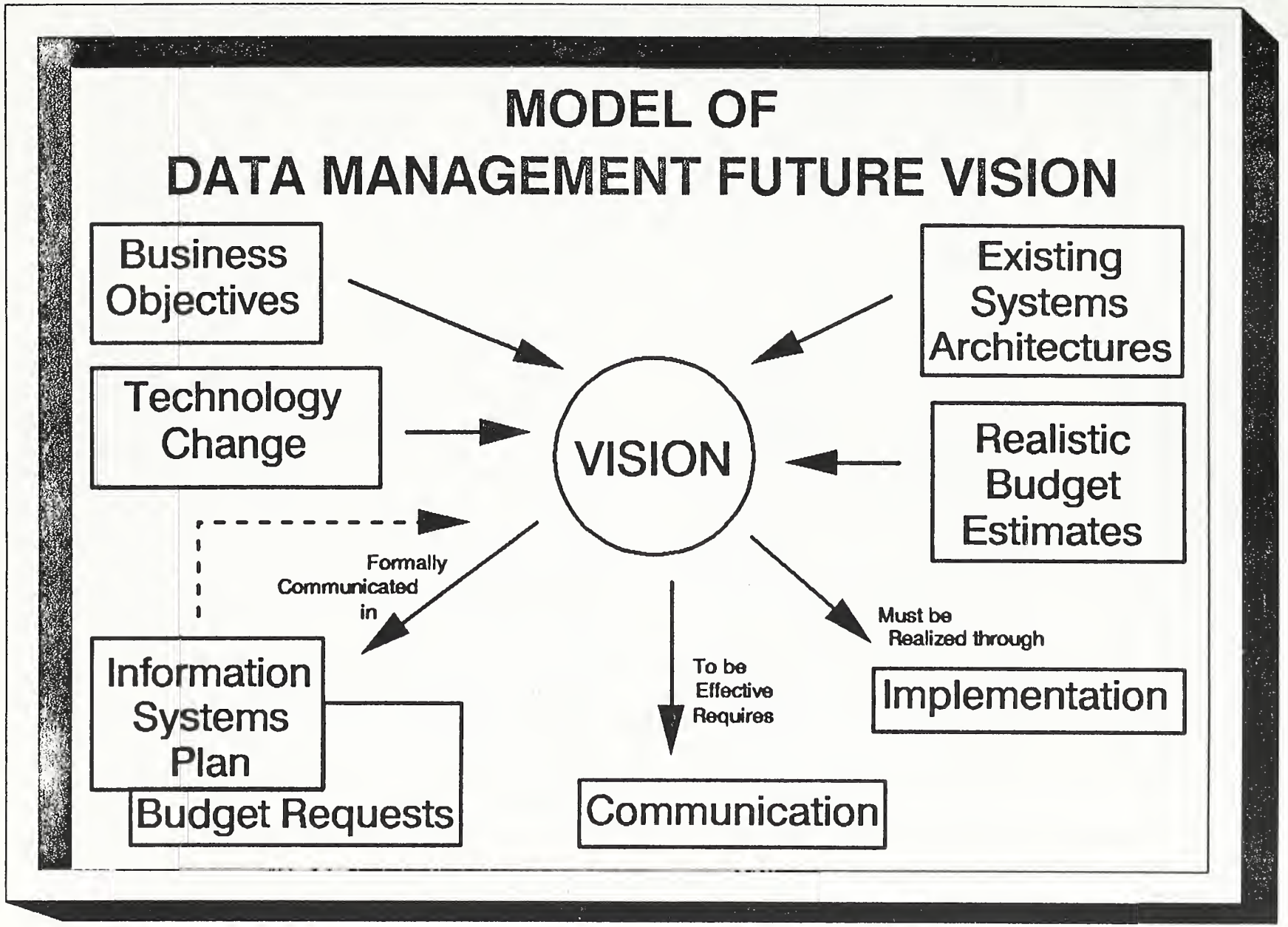
Future vision is a clear, emphatic and powerful statement of what we want our data management program to be. Vision determines our roadmap by defining the starting point and the destination. There is no illusion or daydreaming involved. Nor does it involve an unstructured ad-hoc approach to decision-making where one hopes for the best. Rather, successful architectural change entails knowledge of existing architectures balanced against the unknowns of technological change, current and future business objectives and the realities of budget estimates. Then, our carefully built vision, supported by action, can be a potent tool for focusing the organization towards a common goal.

BUSINESS OBJECTIVES

There is an increasing emphasis on using information systems to support organization objectives, especially in the area of decision-making. It is axiomatic that our future vision must support our organization and its business objectives. However, at times, there is often a tendency in our field to focus on technology capability while disregarding business reality. Decisions based solely on technology consideration must be avoided. In addition to reconciling our vision with business objectives, we must also ensure that the implementation of this vision is cost effective.

TECHNOLOGY CHANGE

Technological advancements came with awesome rapidity in the 1980's and there is every reason to believe this trend will continue and perhaps may even accelerate. What does this mean for our data management vision? It means that a **technology assessment** program is essential to the development of data management vision.
Technology assessment is the identification, understanding and evaluation of new and existing technologies in the context of the organization's goals (business objectives) and needs. Once again,


MODEL OF
DATA MANAGEMENT FUTURE VISION
Business Objectives
Technology Change
Existing Systems Architectures
Realistic Budget Estimates
VISION
Formally Communicated in
Must be Realized through
To be Effective Requires
Information Systems Plan
Budget Requests
Implementation
Communication

this reemphasizes the supporting role of technology vis-a-vis the organization's business objectives. Information sources for technology assessment include literature, conferences, seminars, experts and expert services. The extent of the use of any or all of these information sources depends, in large measure, on the size and complexity of the organization's technology, its mission and available staff and dollar resources. It is necessary both to replace obsolete technology that is no longer cost effective and to avoid the risks associated with unproven leading edge technology. Timing is everything.

EXISTING SYSTEM ARCHITECTURES

Implementation of a data management vision requires a change to system architectures. One of the hardest parts of planning architectural change is figuring out what to discard and what to change.

The existing system architectures describe the information systems environment and serves as a point of departure for envisioning our data management change. In our Commerce Information Resources Management (IRM) Office, we have chosen to describe systems in terms of the architectural structure proposed by Daniel Appleton, namely... **data, applications, control and technology.** In addition, since data management functions span the broad system spectrum of information processing from the collection or acquisition of data to its storage and archival, the development of a realistic future vision is critically dependant upon the comprehensive understanding of the existing architectures.

The **Data Architecture** includes the logical structure which describes the logical data entities and their relationships; the physical structure which is the technical adaptation to a specific system environment; and the user view which is the interface between the individual user and the data. The user view is independent of where or how the data is stored. These data views are interrelated and provide a conceptual bridge which links the manager's or end user's view of the data with the technician's physical view.

Most everyone understands that **Application Architecture** consists of all of the organization's application systems, subsystems and their interrelationships. This architecture is characterized by the presence or absence of uniformity, modularity, reusable standardized components, integration and data sharing. These system development concepts support the current emphasis on port-ability, commonality, and data sharing as being primary areas of system focus. Data-centered application systems, by definition, require increased levels of data management support throughout the system life cycle.

The **Technology Architecture** establishes a hardware and software environment in which the applications and the data reside. The technology base includes **hardware, telecommunications networks, the operating system(s) and the support software.**

The **Control Architecture** defines the precepts and direction governing the other system architectures. It provides a framework of rules and guidelines for the effective management of the organization's information systems processes, operations and support activities and is directed towards the delivery of quality products and services. The Control Architecture is composed of: **standards; policies, regulations and directives; and plans and procedures.**

REALISTIC BUDGET ESTIMATES

A vision which addresses real world considerations cannot be created independent of realistic budget estimates. Budget estimates for data related information systems initiatives are usually based on **cost benefit principles** and business and/or **economic forecasts.** There are some exceptions, such as installing a technical infrastructure in anticipation of potential downstream benefits that are difficult to specify. The objective in developing realistic budget estimates is to minimize risk and ensure that sufficient funding is available to accomplish the task at hand. At times, this is difficult because of the complexity of the systems and technology involved and long lead times to completion which often reduce the accuracy of the estimates and increase the risk of project failure. Budget estimates must be revised when there are changes in the financial forecast and/or the organization's vision which is expressed in the organization's formal Information Systems Plan.

INFORMATION SYSTEMS PLAN

The **Information Systems (IS) Plan**, also known as the Strategic Systems Plan, the Business Systems Plan, the Information Technology Plan, the Information Systems Engineering Plan and by other names, is the means for formally communicating the data management future vision. Although the content and format of the IS Plan varies from organization to organization, there is a core or minimum level of content that exists in most plan versions. The typical core IS Plan includes: the identification and description of planned **information systems initiatives**; the **resources** (men, money and machines) required to support the task effort; and a **milestone schedule for accomplishment** of task objectives to their completion.

COMMUNICATION

It is not enough to merely create the vision, prepare an IS Plan and proceed with implementation. Rather, to be effective, this **vision must be communicated** to and understood by everyone in the organization and ingrained in the organization's operations. If everyone knows the objectives and the plan for getting there, they can move faster and better than without this knowledge. Faulty assumptions, rather than facts, and lack of a common shared vision doom projects to failure while openness and a shared view of the goal significantly increase the chances for success.

IMPLEMENTATION

Implementation is the positive result of moving from a future vision or idea through all of the required phases of development activity to actual installation and/or operation. A vision has little or no worth if it cannot or will not be implemented. Moreover, an important consideration that cannot be ignored involves not losing mission critical functionality and disturbing the normal flow of business when implementing change.

PUBLIC AND PRIVATE SECTOR STATEMENTS

GENERAL

Data collection is now considered straightforward and well established, while the technology for accessing, manipulating and analyzing the data remains complicated and less developed. In fact, it is the very success of some automated data collection efforts that has created the enormous databases that are the source of many data management problems and issues that now require resolution. In response, many information systems managers are redirecting their resources away from the more traditional systems activities and addressing these and other data-related challenges. Typically, the emphasis in many organizations now is on refining the data and delivering it to be shared among users in a manner that supports the organization's mission and provides the capability for better decision-making. Both the public and private sector have responded. It is both informative and illustrative to look at examples of future vision from each of these communities and to compare and contrast their major features.

PUBLIC SECTOR STATEMENT

The Federal Office of Technology Assessment (OTA) issued a report in July 1990 that calls for improvement in data collection, storage, manipulation and dissemination by electronic means. The OTA identified standards for databases, R&D funding for agency information dissemination, sophisticated indexing of databases and documents and increasing the involvement of users in information systems issues as being the requirements needed to achieve successful improvements. The major theme of the report is the critical need to deliver and share important scientific and technical information among those that can benefit. The OTA contends that this is the key to America's competitive position in the world in the next decade.

PRIVATE SECTOR STATEMENT

The main theme presented by John White, of Texas Instruments, and Harry Terrant in a paper titled "Information Management" is very similar to the OTA call for improved data management. White and Terrant propose that there will be a move towards information centered computing where computers will no longer be solely used as electronic

Public Sector Vision
(Office of Technical Assessment)

- Standards for databases
- R & D funding for information dissemination
- Indexing of databases and documents
- Increased user involvement

Private Sector Vision

(White and Terrant - Texas Instruments)

- Sophisticated information retrieval technology
- Software that understands text
- Continuing miniaturization of components
- Awareness of the importance of multimedia

filing cabinets, but will come to be used as intelligent collectors and disseminators of information. They see being able to sort through huge amounts of data to determine what is important and then act on it as the new objective. Sophisticated technology for information retrieval, software that can understand text, continuing miniaturization of components and an awareness of the importance of multimedia applications are required to realize the vision of this new information-centered computing environment.

Improved data management in an information centered computing environment is the goal in both examples. However, they differ as to the requirements to effect the change. White and Terrant focus on improvements and changes to technology, whereas the OTA's vision requires the control (standards) and data (indexing) architectures, budgeting (R&D funding) and end-user involvement (communication) as being the keys for success. Irrespective of differences in their approaches, both parties imply, by their vision, a fundamental change in business systems computing.

DATA MANAGEMENT IN THE DEPARTMENT OF COMMERCE

In the Department of Commerce, the importance of having a data management future vision is a critical success factor for the successful accomplishment of our major programs. The Bureau of the Census, the Patent and Trademark Office (PTO), the National Technical Information Service (NTIS), and the National Oceanic and Atmospheric Administration (NOAA) with it's Weather Service, Environmental Satellite Service, Ocean Survey, Fisheries, and Oceanic and Atmospheric Research components, process and generate massive amounts of information. Given their size, the National Institute of Standards and Technology (NIST), the Bureau of Economic Analysis (BEA), the Bureau of Export Administration (BXA), the Economic Development Administration (EDA), the International Trade Administration (ITA) and the Office of the Secretary also manage large databases.

Commerce bureaus have considerable diversity with respect to their mission programs and, consequently, have differing data management problems, requirements and visions. Regrettably, time does not permit a review of each bureau's circumstances and needs. Therefore, NOAA, Census, PTO, NIST and the Office of the Secretary have been selected for further discussion on this subject. The first three bureaus are included because of the size and scope of their data management activities and because their differing data related missions and problems are representative of those found in the other bureaus. NIST has an important and unique national mission that supports the effective management of data. The Office of the Secretary's current data management initiatives are also presented together with a global vision of data management in Commerce by the end of the decade.

NOAA

The National Oceanic and Atmospheric Administration has massive data-related problems in carrying out its environmental program missions which are interagency, national and international in scope. NOAA now acquires about 2 terabytes of data every year and its data management systems are operating at full capacity. In fact, NOAA has worked hard to keep pace with the 10% annual growth the last several years. It is estimated that NOAA will have to handle 200 terabytes per year by the year 2000. In response, NOAA has initiated the Earth System Data and Information Management (ESDIM) Program as a statement of their vision to meet this challenge.

The ESDIM Program is a NOAA-wide approach to integrate common elements of data and information management through technology information. The centerpiece of the program is a distributed data network that will feature several main environmental data centers or libraries that will evolve from current NOAA Data Centers. The main data centers will be connected to many semi-autonomous nodes which will contain the data supporting specific NOAA line organizations and programs. ESDIM will be controlled and managed by a master directory which will contain all of the data and information holdings in the network and the directions for their access. This program is a long-term effort that will affect almost every aspect of NOAA's activities and will also require active participation from external communities. Clearly, NOAA has formed their future vision and is making impressive progress towards its realization.

CENSUS

While the Census Bureau must contend with an expanding demographic database that is the result of population growth and new data, their primary data management concern is the timely and accurate capture and processing of census respondent entries. During the 1990 Decennial Census, more than one hundred million questionnaires were processed through the FACT 90 (FOSDIC and Automated Camera Technology 1990) during the first one hundred days. This advanced technology, which was engineered to capture data from questionnaires under uncontrolled conditions, employs high resolution cameras for microfilming the questionnaires and computer scanning to capture the data entries. Advanced document handling technology and controlling the paper characteristics and print quality were essential to the success of the data capture system.

The Census Bureau has begun active planning for the 2000 Decennial Census. This early start provides an opportunity to identify and develop new ways to design and implement this massive data collection, processing and dissemination effort. The amount of information necessary to administer a huge temporary workforce, assure the quality of response by all residents of the Nation, and capture and tabulate the results in a timely manner is astounding.

For the 2000 census, their proposed objective is a totally integrated system that would contain and manipulate the administrative and substantive data needed to control and complete the census. Control files based on persons and the housing inventory will provide the capability to attach progress, personnel, and payroll information to generate records for managing hundreds of thousands of staff and the millions of data items collected. Interactive control file searches would allow immediate linkage between addresses, persons, and census data to be captured from a variety of input sources, edited and coded immediately, and tabulated efficiently at a rate not allowed by current systems. Driven by the operational success of the 1990 census, the Census Bureau has developed a future vision to work towards as they approach the 2000 Decennial Census.

PTO

The U.S. Patent and Trademark Office is building an Automated Patent System (APS) that will give patent examiners and the public the capability to search, retrieve, and print patents, both U.S. and foreign, in an automated environment, requiring little support from paper files. The data requirements for the system are enormous. At present, there are approximately one million patents loaded as images (raster scanned data) to optical disk storage platters. The combined size of the current image database is about 150 gigabytes. In addition, the PTO has a U.S. patent full text database that is fully word searchable for nearly all patents issued from 1971 to the present. The current size of the file is approximately 6 gigabytes with about 100 megabytes added each issue. The amount of data stored in the APS presents many challenges to the PTO data administration/management staff both now and in the future.

By 1999, the PTO plans to have 6.7 million U.S. patents, 8 million Japanese patents and 6.3 million European patents online, stored on optical platters for retrieval. The amount of image data is expected to equal about 27 terabytes. The most formidable challenge for the future is ensuring high data integrity. Their objective is to acquire reliable software that will perform much of this effort in the future. However, it is not certain how much can be automated. By way of support for these systems objectives, PTO is developing an aggressive data management program and will use every tool available to ensure the integrity of its data.

NIST

The National Institute of Standards and Technology's Computer Systems Laboratory (CSL), under the leadership of James Burrows, has an altogether different data management challenge since the Laboratory has the important national mission of developing computer and related telecommunications standards (Federal Information Processing Standards--FIPS), policies and guidelines and providing technical assistance to federal agencies. CSL has another national role as a primary contributor to the development of American National Standards

Institute (ANSI) standards. On the international level, CSL is a leader in the International Standards Organization (ISO).
As the pace of information systems technology growth continues to accelerate and become increasingly more complex, the pressure for more standards to support management of the information systems environment will also increase. This scenario presents a problem of sizeable proportion both for ANSI and CSL management. However, much of the progress we will achieve in realizing our data management vision will depend on the CSL's continued efforts and quality leadership in the standards area.

OFFICE OF THE SECRETARY

In the Office of the Secretary, we've recently established a Data Management Division under IRM for the purpose of ensuring the effective administration, integration and delivery of our Commerce administrative and management information. The Data Management Division also develops and manages the integrated Management Information System. In addition, we have established a Commerce Data Management Interest Group in order to benefit from the extensive knowledge and experience of our bureau experts in data management and information systems. The purposes of this Group are to: promote and facilitate the transfer of data management information between Commerce organizations; share resources; and serve as a technical management forum to provide input to the formulation of Commerce Data Management policy.

DATA MANAGEMENT IN COMMERCE IN THE YEAR 2000

Our global vision of data management for Commerce in the year 2000 encompasses Secretary Mosbacher's goal, our IRM objectives and critical bureau mission needs, together with predictions of future technology. It includes:

- Powerful mainframe computers in the 100 plus MIP (millions of instructions per second) range that are used for scientific "number crunching" and as intelligent centers for data warehousing and distribution. Far from becoming obsolete for business processing, these centers will store and process enormous databases, perform image processing operations and also serve as network managers for sophisticated distributed data networks. Their sophisticated security mechanisms will provide a more secure environment for accessing data resources and for data transmission than will exist in Local Area Networks (LANs) and individual workstations. In addition, the mainframe backup capabilities for data and operations will remain superior to that of other distributed networks.

- Telecommunications improvements which feature fiber optics and high performance communications will support the accessing and transmission of massive amounts of data. Distributed data complexes will be coupled by wide band, wide area communications networks. Each complex, while self-contained, will require information sharing

Commerce in the Year 2000
Global Vision includes:
Powerful mainframe computers
Telecommunications improvements
Optical and magnetic media storage
Sophisticated software for analysis
New and enhanced data standards
New organizational roles

with other nodes and the information processing center. The implementation of the Open Systems Interconnection standard, the proliferation of LAN's, and the remote storage of large amounts of data will create the requirement for reliable, high quality communications.

- Optical and magnetic media storage technologies which will greatly increase our capabilities for data capture and storage. Low cost imaging will have proliferated throughout the work place. In fact, imaging and CD-ROM, together with the maturation of now leading edge technologies such as voice automation and handwritten text input, will provide the basis for another evolutionary step in office automation. CD-ROM in its many forms, including integrated multimedia (audio, video and digital data), will have a paramount place in this office revolution.

- Sophisticated software with an expert systems' core will automate the analysis of data. Automated analysis software will narrow the gap between the collection of data and refining it into wisdom or information by applying intelligence. In addition, software designed specifically for the management of data will increasingly assume such tasks as automatic backup operations, archiving, dissemination and conversion.

- New and enhanced data standards and standards for other technologies which will have been implemented. These standards will have eliminated the potential for variability and resulting chaos that might otherwise occur. Standards that support the effective management of data will be found in the areas of: data analysis and design, data security, data modeling, data formatting and content, data transmission and interchange, document management and design, imaging, data description, data storage and databases.

- New organizational roles for the data management function and for the information systems professionals having knowledge and proficiency in this area of expertise. Data management will be more specialized and the information systems professionals that support the function will move higher in the organization to act more as enterprise consultants. Technical training, systems problem solving, operations and other technical matters will continue to be handled at lower levels in the organization.

This vision may seem to be beyond our reach. However, given the continuing strong support from the executive and legislative branches; the commitment, professionalism and enthusiasm of our bureau program managers and their technical staffs; the support and encouragement of our IRM units; and more rapid progress in developing information systems technology, we'll be a long way towards realizing the success of our vision.

NIST-114A
(REV. 3-90)

U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION OR REPORT NUMBER
NISTIR 4843

2. PERFORMING ORGANIZATION REPORT NUMBER

3. PUBLICATION DATE
MAY 1992

4. TITLE AND SUBTITLE
Managing Data - From Vision to Reality
Proceedings of the Fourth Annual DAMA Symposium

5. AUTHOR(S)
Judith J. Newton, Mary Lou Melley, Holly Higgins, Editors.

6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS)
U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
GAITHERSBURG, MD 20899

7. CONTRACT/GRANT NUMBER

8. TYPE OF REPORT AND PERIOD COVERED

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)
National Capital Region Data Administration Management Association
PO Box 598
Columbia, MD 21045-0598
and NIST

10. SUPPLEMENTARY NOTES

11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.)

The National Capital Region chapter of the Data Administration Management Association (DAMA) sponsored its fourth annual symposium on May 14-15, 1991. The setting was the National Institute of Standards and Technology in Gaithersburg, Maryland. Several hundred attendees from the public and private sectors were treated to the wisdom of data administration practitioners and preachers at the leading edge. The topic of the 1991 symposium, "Managing Data: From Vision to Reality," set the framework for an enriching two days. The papers in this proceedings highlight the actual presentations.

12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)

data administration; data management; information resource management.

13. AVAILABILITY

[X] UNLIMITED

[ ] FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS).

[ ] ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402.

[X] ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.

14. NUMBER OF PRINTED PAGES
166

15. PRICE
A08

ELECTRONIC FORM